## AD NUMBER

**ADB216627**

## NEW LIMITATION CHANGE

TO

**Approved for public release, distribution unlimited**

FROM

**Distribution authorized to U.S. Gov't. agencies only; Specific Authority; Oct 96. Other requests shall be referred to Commander, Army Medical Research and Materiel Command, Attn: MCMR-RMI-S, Fort Detrick, Frederick, MD 21702-5012.**

## AUTHORITY

**USAMRMC ltr, 4 Dec 2002**

AD_____

CONTRACT NUMBER DAMD17-96-C-6022

TITLE: Neural Network Medical Decision Algorithms for Pre-Hospital Trauma Care

PRINCIPAL INVESTIGATOR: B. Eugene Parker, Jr., Ph.D.

CONTRACTING ORGANIZATION: Barron Associates, Inc.
Charlottesville, VA 22901-1444

REPORT DATE: October 1996

TYPE OF REPORT: Final, Phase I

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | October 1996 | Final, Phase I (15 Mar 96 - 14 Sep 96) |

**4. TITLE AND SUBTITLE** Neural Network Medical Decision Algorithms for Pre-Hospital Trauma Care

**5. FUNDING NUMBERS**
DAMD17-96-C-6022

**6. AUTHOR(S)**
B. Eugene Parker, Jr., Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Barron Associates, Inc.
Charlottesville, VA 22901-1444

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, Maryland 21702-5012

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Distribution authorized to DOD components only (specific authority, Oct 96). Other requests for this document shall be referred to Commander, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-RMI-S, Fort Detrick, Frederick, MD 21702-5012.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200*

This SBIR Phase I research project is concerned with the problem of civilian and military trauma management, whose paramount issues include: (1) obtaining knowledge about the physiological condition of the injured patient (e.g., injury severity assessment and survival likelihood prediction); and (2) making intelligent use of that information for pragmatic decisional purposes (e.g., triage).

The emphasis of our research effort is on assessing the ability of polynomial neural network (PNN) methods to improve on conventional trauma scoring systems and other modeling approaches, such as logistic regression. Using several real-world civilian trauma registry databases, we demonstrated: (1) that PNN models can provide significant improvement over existing pre-hospital and ex post scoring systems, such as T-RTS, TRISS, and ASCOT, in terms of the specificity-sensitivity characteristics of mortality prediction; (2) the ability to discriminate accurately among three or more classes of patients (e.g., RED, AMBER, and GREEN triage categories); (3) the ability to compensate for missing input variables while achieving results not significantly different from those obtained using models that did not rely on such inputs; and (4) the ability to obtain superior performance through time-series modeling of available patient data.

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Trauma scores, Triage, Estimation, Classification, Polynominal neural networks, Statistical induction; Pre-hospital care; Post-hospital assessment | 89 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Limited |

## Foreword

Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

BEP Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA, the investigator(s) adhered to current Guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide of Biosafety in Microbiological and Biomedical Laboratories.

B Eugene M 10/12/96

PI - Signature Date

## Acknowledgments

This document is the Final Technical Report for Contract DAMD17-96-C-6022, an SBIR Phase I Award funded by the U.S. Army for the six-month period from March 15 to September 14, 1996. The research effort, which was based on the July 7, 1995 DoD SBIR Program Solicitation 95.3 Topic A95-095 entitled "Medical Decision Algorithms for Pre-Hospital Trauma Care," was carried out by Barron Associates, Inc. (BAI) of Charlottesville, Virginia. Under the direction of the project Principal Investigator, B. Eugene Parker, Jr., Ph.D., Edward C. Larson of the BAI technical staff conducted most of the data analyses documented herein and was a primary contributor to this report.

The authors express their appreciation to Major Stephen P. Bruttig, the Contracting Officer's Representative, for his helpful guidance and encouragement of BAI's efforts. The collaboration of George H. Lindbeck, M.D., of the University of Virginia Hospital, Division of Emergency Medicine, who served as a consultant in this project and who provided the UVA trauma data, is gratefully acknowledged. We also thank Robert Rutledge, M.D., of the University of North Carolina at Chapel Hill School of Medicine, Department of Surgery, who provided the North Carolina Intensive Care Unit and Trauma Registry databases, along with valuable critique of our work.

# Contents

# List of Figures



# List of Tables

# Abbreviations and Acronyms

| | |
|---|---|
| AGE | Patient's age |
| AI | Artificial Intelligence |
| AIS | Abbreviated Injury Scale |
| APACHE | Acute Physiology and Chronic Health Evaluation |
| ASCOT | A Severity Characterization of Trauma |
| BAI | Barron Associates, Inc. |
| B/P | Blunt/Penetrating |
| CNS | Central Nervous System |
| CO | Cardiac Output |
| COV | Coefficient of Variation |
| CRAMS | Circulation, Respiration, Abdomen, Motor, Speech |
| CPP | Cerebral Perfusion Pressure |
| DBP | Diastolic Blood Pressure |
| EMT | Emergency Medical Technician |
| EY | Eye Component of Glasgow Coma Scale |
| ER | Emergency Room |
| FIR | Finite Impulse Response |
| FSE | Fitted Squared Error |
| GCS | Glasgow Coma Scale |
| GMDH | Group Method of Data Handling |
| HCT | Hematocrit |
| HR | Heart Rate |
| ICD-9 | International Classification of Diseases, 1989 |
| ICP | Intracranial Pressure |
| ICU | Intensive Care Unit |
| ISS | Injury Severity Score |
| KRC | Kane's Revised Checklist |
| MLP | Multilayer Perceptron |
| MOI | Mechanism of Injury |
| MT | Motor Component of Glasgow Coma Scale |
| MTOS | North American Major Trauma Outcome Study |
| NCICU | ICU database from University of North Carolina |
| NCTR | North Carolina state Trauma Registry |
| PHI | Pre-Hospital Index |
| PNN | Polynomial Neural Network |
| PSE | Predicted Squared Error |
| RMS | Root-Mean Square |
| RNN | Recurrent Neural Network |
| ROC | Receiver-Operator Characteristic |
| RPM | RR, Pulse (heart rate), and MT Score |
| RR | Respiratory Rate |
| RSG | RR, SBP, and GCS Score |

| | |
|---|---|
| RSM | RR, SBP, and MT Score |
| RTS | Revised Trauma Score |
| SAPS | Simplified Acute Physiology Score |
| SBP | Systolic Blood Pressure |
| SEV | Paramedic Gestalt Impression |
| T | Body Temperature |
| TRISS | Trauma and Injury Severity Score |
| T-RTS | Trauma Score for Triage |
| TS | Original Champion Trauma Score |
| TTR | Baxt Trauma Triage Rule |
| UVA | Trauma database from University of Virginia Hospital |
| VB | Verbal Component of Glasgow Coma Scale |
| WDMEV | Wound Data and Munitions Effectiveness in Vietnam |

# 1 Goals and Scope of Project

The following description of the goals of this project, and upon which all of the research endeavors herein are based, is verbatim from the 1995 Army SBIR Solicitation (Topic A95-095):

**OBJECTIVE:** To develop computer algorithm(s), capable of accepting data from physiological sensors already under development, which will operate in small, hand-held personal computers such as the Soldier Individual Computer, 21st Century Land Warrior (adapted for medical applications).

**DESCRIPTION:** This decision algorithm must be capable of accepting multiple inputs, (such as tissue pH, tissue $O_2$, tissue blood flow, cardiac output, heart rate, ambient temperature, and body temperature), and provide output in 15 seconds or less. Output would be a combination of "likely survival" and "approximate survival time," which could each be digitally displayed, but must be displayed as RED, AMBER, GREEN (RED=death imminent; physiological and physical parameters 20% of "normal;" AMBER=serious to extraordinary deviation from normal physiology – death likely in 30-60 minutes; physiological and physical parameters 50% of "normal;" GREEN=survival likely; physiological and physical parameters within 80-100% of "normal.")

**PHASE I:** Develop realistic algorithms based on scientific literature values, previous models and validated assumptions, including descriptions above.

**PHASE II:** Validate algorithm with experimental data; refine algorithm, compile algorithm and necessary supporting software, drivers, etc. for incorporation on microprocessor chip. Phase II model must be capable of updating data from previous readings, in order to determine whether intervening treatment was effective, or whether spontaneous course of casualty is changing.

The SBIR Phase I research project that we have carried out is concerned, at the highest level, with the problem of *trauma management*, whose main goal is to minimize loss of life from traumatic injuries sustained by human beings, e.g., combat soldiers. The most paramount issues in critical care medicine, in both civilian and military contexts, pertain fundamentally to a key two-stage process, namely: (1) obtaining knowledge about the physiological condition of the injured patient (e.g., injury severity assessment and survival likelihood prediction); and (2) making intelligent use of that information for pragmatic decisional purposes (e.g., triage).

To both of these challenging tasks, Barron Associates, Inc. (BAI) offers distinguished acumen and specialized expertise in mathematical methodologies and software tools, most notably *polynomial neural network* (PNN) synthesis algorithms. The present Final Technical Report focuses on the application of PNNs and other phenomenological modeling methods for optimal use of pre-hospital medical data. We explore the distinctive strengths and capabilities of neural network models derived from empirical databases of historical trauma cases. Estimation and classification models for injury severity assessment and survival outcome prediction are developed and compared to conventional scoring systems for both pre-hospital and *ex post* trauma evaluation. The main goal of our research efforts, pursuant to the stated goals of the solicitation request, is to assess the benefits that PNN methods can bring to the field of trauma management and their ability to achieve superior performance over conventional scoring systems and other modeling tools such as conventional logistic regression.

# 2 The Trauma Care Environment

## 2.1 Definition of Trauma

*Trauma* encompasses a wide variety of injury types and causative mechanisms (e.g., blunt head concussion from fall or automobile accident, drowning or ingestion of poison, penetrating bullet or knife wounds, burns, etc.), all of which do harm in at least one of two ways: (1) bodily tissue is anatomically ruptured or otherwise damaged by destructive energy from the external environment; and (2) normal physiological function is disrupted or endangered. Depending on its exact nature, a traumatic event results in a complex state of physiological disturbance, the severity of which may range from mild to immediately life-threatening. Medical intervention is often necessary to make the difference between life and death. Trauma, which accounts for most instances of what is commonly deemed "unnatural" death, is the third leading cause of death in the United States, behind cancer and cardiovascular disease, with automobile accidents alone taking several hundred thousand lives per year [43]. In the first four decades of life, trauma is the leading cause of death and accounts for the majority of pediatric deaths.

## 2.2 Civilian Trauma Management

To initiate discussion of trauma management, it is useful and illuminating to discuss its practice in the civilian world, even though the main interests of the solicitation topic are military-oriented. It is, after all, in the civilian realm that trauma management is practiced on a routine, ongoing basis. Standards of care are high, and support infrastructure (e.g., large emergency departments in urban hospitals, well-maintained fleets of evacuation ambulances and helicopters, first-rate teams of experienced paramedics regularly on call) is highly developed and well-financed in most parts of the country. Moreover, almost all of the existing body of research in trauma management to date, such as the North American Major Trauma Outcome Study (MTOS) [19], has been exclusively in the civilian realm. The preponderance of injuries encountered in such studies are blunt trauma to the head, spine, or thorax resulting from automobile collisions. Most of the trauma evaluation methodologies, or scoring systems, emerging from such research were designed primarily for quality-of-care control and comparison by providing benchmarks for trauma management practices utilized by different hospitals.

Trauma management involves several key elements and themes: the injured patient, emergency medical technicians (EMTs), medical techniques and protocols for diagnosing the patient, treatment options, evacuation modalities, communication, and hospital facilities. Following any traumatic event, whose occurrence is always unexpected and random, pre-hospital trauma rescue involves arrival of EMTs at the scene of injury, administration of first-aid treatment, evacuation from the scene of injury, and triaging the patient to an appropriate hospital facility. There is enormous variation in how this sequence of operations is implemented in local communities, which differ demographically, topographically, and in the infrastructure capacity and quality of services they can provide.

To prevent loss of life, all of these steps must be performed skillfully within the well-recognized "golden hour" [74] after the traumatic event. The guiding objective of the EMT team throughout the rescue process is to take action deemed necessary to preserve the life of the patient at hand and to triage that individual to the nearest hospital facility best equipped to deliver the needed intensity of care. The philosophy is conservative in that the penalty of undertriage (losing a patient who

"could have been" saved) is immeasurably greater than that of overtriage (sending a noncritical patient to a hospital facility intended for only the most serious cases). Even though the excess capacity that major hospitals must carry to cover such cases may justifiably be regarded as a small price for saving lives, the burdens shouldered by emergency departments throughout the country in both rural and urban areas, and the concomitant jeopardy to other patients due to delays, should not be underestimated. Often, the onus of high arrival rates is excessive, straining precious critical care resources beyond the utilization levels at which they can operate most efficiently and effectively. The causes of such overload [65] are primarily sociological, e.g., escalating homicide rates, mounting congestion and long commutes on roads, growing demand for around-the-clock availability of medical care, tendencies of busy or off-duty private physicians to refer patients to emergency departments, and health insurance policies that require initial emergency room supervision for coverage of ensuing long-term rehabilitation costs. Even though these issues *per se* are not part of trauma management (they merely make treatment resources effectively more scarce than they would otherwise be), it is nevertheless true that better pre-hospital *triage*, i.e., referral decisions by EMTs, would help reduce emergency department overload significantly. Accurate identification of pre-hospital patients *not* in need of critical care facilities, while keeping undertriage rates acceptably low, beckons a role for diagnostic procedures and forecasting tools for use by pre-hospital EMTs in the field. Emergency room triage nurses may also find such tools to be valuable and helpful. Use of such information-processing algorithms for diagnostic, prognostic, and decisional purposes could help minimize loss of life.

## 2.3 Military Trauma Management

Trauma management in the military realm is profoundly different from that in the civilian world in several major respects. First, the lethality of the environment is incomparably greater. Depending on the nature of the fighting, as many as 80% of combat injuries may be fatal, with most deaths occurring before any medical assistance at all can arrive. 40% of fatally wounded soldiers die within 15-20 minutes, and 70% die within one hour. Injuries encountered in combat are typically extremely severe; more than 90% stem from bullet or shrapnel penetration. Blast and thermochemical injuries account for most other casualty incidents. Among penetrating injuries, the great majority are wounds to the skull, heart, or great vessels. Tissue destruction is especially severe for high-velocity wounds in which the impinging shell is designed to explode or tumble, thereby dispersing large amounts of bulk tissue and creating large exit wounds. These types of injuries are often gravely underestimated by such conventional indicators as the Abbreviated Injury Scale (AIS) and Injury Severity Score (ISS), which are used primarily for post-surgical evaluation of civilian incidents. Massive exsanguination, sepsis, and arrest of central nervous function pose threats to life so grave that the golden hour paradigm ceases to be relevant. The urgency of most military trauma cases is better summarized as the "brass ten minutes" [12].

The lethality of the ground combat environment is often so great that rescue personnel risk their own lives in attempting to reach wound victims. Rescue should be attempted, therefore, only if it stands to make a difference between life and death for the wounded and the rescuer has a good chance of reaching him and escaping unscathed. It is hard to imagine any such comparably extreme conditions in civilian life apart from riots or violent demonstrations. Attempting to access accident victims is usually itself safe, with exceptions such as extricating victims from damaged buildings or frigid water, and is seldom a matter of deliberation at all.

To make matters more grim, treatment resources during war are almost always in scant supply and are often primitive by civilian standards. It also requires considerably more time, risk, and effort to transport a wound victim to the equivalent of a full-service hospital during war than it does in typical peacetime scenarios. In the military setting, battalion aid stations or similar sites are often the first destinations to which wounded soldiers are transported. Among those who do reach the hospital alive, however, a large majority survive [12]. The unsafe and chaotic nature of the battle environment sometimes makes adherence to established treatment and evacuation protocols prohibitively difficult. Battalion groups often have little choice but to rely on makeshift teams of rescue personnel who are no match for civilian EMTs in medical expertise or knowledge. Many such rescue workers do not even have the first-aid training to infer such basic indicators as respiratory status, heart rate, and level of consciousness. For this reason, there is a niche for decisional algorithms that would *instruct* such personnel as they tend to a patient. Many acute battlefield injuries, such as tension pneumothorax (see Appendix B), often remain undiagnosed until it is too late, but treatment as crude as piercing the chest could save the victim's life.

In conjunction with specialized biomedical instrumentation to automate diagnostic procedures, computer-driven algorithms would be immensely helpful in the information processing aspects of pre-hospital triage, i.e., in diagnosing the patient and making the best use of that information. In both realms, there is thus a compelling demand for advanced information processing capability in the field. Whereas the added value in the civilian realm would be primarily in risk stratification of incoming patients to a hospital, the role in the military would primarily be for triage prioritization. Both realms could, however, benefit from the application of standardized medical protocols (e.g., a "doctor-in-a-box") to guide treatment.

# 3 Algorithms for Information Processing and Decision Support

The purpose of on-line algorithms is to obtain *information*, as opposed to merely a collection of facts, on the spot. In pre-hospital trauma care, as in an enormous number of other applications, it is incumbent upon the analyst (in this case, the EMT or physician) to contend with a sizable body of *data* and transform it into pragmatically useful and relevant information, upon which critical decisions rely. Factual data, as they are acquired, very often impinge on the human mind as being superficially disparate and bewildering. Often, salient data *are* self-conflicting and, to make matters worse, may not even be accurate due to limitations and biases in data acquisition processes (including human intuition). The psychocognitive difficulty of the human analyst's grappling with a daunting array of facts is compounded all the more under the duress in which pre-hospital trauma rescue must be performed.

The problem of distilling a concise and manageable kernel of information, directly useful for decisional purposes, from raw data is known as *information processing*. Several very general approaches exist, the least sophisticated (and arguably the most commonly used in the real world of human affairs) is seat-of-the-pants intuition and instinct. It relies in large part on "common sense" and anecdotal knowledge on the part of the practitioner. Although the efficacy of sheer intuition should not be dismissed out of hand, it is certainly fair to ask whether more systematic, less subjective inference and decisional methods can improve trauma management. Conventional trauma management protocols jointly employ two approaches that are a step above raw intuition in sophistication: (1) EMT judgment; and (2) rote scoring systems. The former is based on extensive clinical experience in the field and academic medical training. Although it is often very effective,

sometimes impressively so, it requires an allocable supply of professionally trained emergency workers who, in practice, are often not available. In both civilian and military settings, the scarcity problem would be alleviated greatly if it were feasible for less experienced personnel to undertake the responsibilities of pre-hospital trauma treatment, evacuation, and referral.

## 3.1 Strengths of Algorithmic Approaches

A large number of rote scoring systems have been introduced by various researchers, a representative sampling of which are described in detail in Appendix A. These systems appeal to a basic common set of physiological, neurological, anatomic, and cause-of-injury criteria to evaluate the medical condition of a trauma victim and the immediacy of required hospital care. These protocols are typically formulated as formal checklist or flow-chart procedures, boiled down to essentially an index-card minimum of steps that an emergency field worker, with just a smattering of training and modicum of practice, should be able to memorize and perform routinely. Many experienced, well-trained EMTs often consult such rule-of-thumb procedures in their own evaluations of patients. Because the steps must be performed mentally without cue cards, the rules must be kept extremely simple, containing five or fewer steps. As arithmetical computations must be kept to an absolute minimum, the data variables (e.g., SBP, RR) are almost always coded, i.e., converted into small numbers: $0, 1, 2, \ldots, 6$. Although some quantitative information is discarded by this process, the utility (and ingenuity) of coding should not be underestimated, since the range that a given coded value covers may coincide with a certain *qualitative* physiological state, knowledge of which could have direct bearing on treatment and triage decisions. Stated alternatively, there may be fairly sharp boundaries separating normal from abnormal physiology that shrewd placement of coding cutoffs might manage to capture. It is in part this fundamental notion, in fact, that justifies a key role for *classification* methods in trauma management.

Nevertheless, both EMT judgment and scoring systems are highly fallible predictors of injury severity. A problem common to many scoring systems, for instance, is that they fail to achieve simultaneously high *sensitivity* (identification of patients requiring intensive trauma care) and *specificity* (identification of patients *not* requiring intensive trauma care). In the interests of conservatism, one generally opts for the former at the expense of the latter; in other words, tolerating some overtriage to keep undertriage minimal. The underlying difficulty appears to be not so much an information processing problem, i.e., what is done with data once it is acquired, but rather, inherent limitations in the data set itself. In other words, the problem is one of *observability*: the data fields (GCS, SBP, RR, HR, B/P, and AGE) common to most of the conventional scoring systems seldom capture the true nature of physiological and neurological disturbance sustained by the trauma patient to the degree necessary for truly effective trauma management. For this reason, even the most sophisticated and well-honed quantitative algorithms synthesized using extensive, high-quality databases are inherently limited in the prediction performance achievable. Nonetheless, it is still necessary and worthwhile to explore algorithmic approaches to trauma management for at least the following reasons:

- They can potentially provide improvement over conventional methods, which are deliberately minimal in the complexity of the steps they employ.
- Computerized decisional support instrumentation will be needed to provide directions and assistance to inexperienced personnel performing emergency trauma treatment during war.
- *Algorithms*, which, by definition, are unambiguous procedures, are invaluable from a qual-

ity control and database construction perspective in that they reduce reliance on subjective judgment and help standardize definitions and conventions among various institutions and researchers.

- Algorithms are inherently faster and more consistent than human thought processes.
- As biomedical technology capable of acquiring comprehensive, high-quality, physiological data (e.g., blood gas profiles, hemodynamics) at the scene of injury becomes available, algorithmic methods will be necessary to draw meaningful conclusions from such data.
- As trauma simulation capabilities increase, algorithmic methods will still be necessary to draw meaningful conclusions from such simulations.

## 3.2 Inference and Prediction

There are two general distinctions in algorithmic applications that are noteworthy in the present context, the first of which is between *inference* and *prediction.* Both seek to produce useful information on the spot from a jumbled collection of data. However, the former attempts to extract factual information about the past or present, whereas the latter, by contrast, seeks to predict what might happen in the future. A tool to infer ISS (Appendix A.1), for instance, based on physiological data and anatomical evidence acquirable at the scene of injury, is an example of an inference algorithm. It discerns, philosophically speaking, the objective reality of an event that has already taken place. Inference algorithms are intended to serve as "virtual sensors," i.e., substitutes for tangible instrumentation capable of measuring the quantity of interest directly. ISS, in this case, is meant to describe the severity of an injury that has already occurred. Present biomedical technology does not permit the EMT to "see" the injuries inside the patient that contribute to the constituent AIS scores, but a virtual sensor would perform that function indirectly by utilizing physically observable indicators that correlate (linearly or nonlinearly) with ISS. However, to the extent that these correlations are uncertain or that information about the past or present is incomplete, the inference process is stochastic.

By contrast, prediction algorithms are concerned with events that have not yet occurred. For example, a quantitative method to determine the survival chances of a critically injured patient, assuming administration of a particular conditional treatment regimen, is an example of a prediction algorithm. Stated in alternative language, inference algorithms perform *diagnostic* functions, whereas prediction algorithms perform *prognostic* functions. The latter require special consideration because they require not only (complete or incomplete) knowledge of the present state of the system, but also explicit assumptions about what will happen in the future. For example, predicting how long a critical patient is likely to survive is a prognosis problem extremely pertinent to battlefield triage. Survival time projections are contingency predictions with respect to a repertoire of alternative treatment policies that *could* be followed.

## 3.3 Estimation, Classification, and Decisional Algorithms

The second major distinction in information-processing algorithm applications is between *estimation* and *classification*, the output types of which are fundamentally different mathematical entities. Estimators generate output information in the form of a numerical quantity, usually a real number, that can assume a continuum of possible values. An algorithm that forecasts, say, the approximate survival time of a critical wound victim is an example of an estimation algorithm for

prediction, since time is a continuous variable. The simplest and best-known mathematical procedure for constructing phenomenological estimation models from empirical data is least-squares regression.

By contrast, classifiers generate discrete categorical outputs. In classification problems, one seeks to sort individual cases, or *exemplars*, into one of two or more classes based on certain observable attributes they exhibit. Sorting trauma victims into color-coded triage groups, based on such readily observable attributes as GCS and RR, is a prime example. Numerous other illustrations arise in epidemiology, such as predicting the odds of breast cancer or coronary heart disease based on medical risk factors and demographical profiles for various individuals. In such applications, the outputs represent probabilities of membership in various non-numerical categories (e.g., POSITIVE/NEGATIVE, TRUE/FALSE, LIFE/DEATH, RED/AMBER/GREEN, etc.). The outputs, in other words, are merely labels, which may be unordered (known as nominal-level class sets) or ordered (ordinal class sets). In nominal-level cases, the categories have no implied rank ordering or relationship to one another (e.g., an anatomical inference classifier that identifies internally hemorrhaging organs and produces outputs LIVER, LUNG, or SPLEEN). In both nominal and ordinal cases, the space of output classes is non-metric in that no well-defined measure of distance between any two classes is recognized. For example, in the RED/AMBER/GREEN problem, which we treat in detail later on, REDs are defined as pre-hospital fatalities, GREENs as those highly likely to survive, and AMBERs as those critical cases in which survival is uncertain. It makes sense to say that AMBER lies *between* RED and GREEN, but it is not admissible to recast it as an estimation problem, with RED = 3, AMBER = 2, and GREEN = 1. This would fallaciously imply that AMBER lies exactly halfway between RED and GREEN, whereas the labels describe broadbrush categories that have certain probability distributions over attribute space and may overlap (see Appendix C.1).

This type of subtlety makes classification fundamentally different from estimation. The mathematical methods for fitting classification models to empirical data, moreover, are markedly different from those for constructing estimation models. Whereas least-squares regression is the basic method for deriving estimation models, the analogous technique for deriving classification models is *logistic regression*, the mechanics of which are presented in Appendix C.3. The outputs of classification models are *probabilities* that a given exemplar belongs to each possible class, such that the membership probabilities summed over all classes in the output space sum to unity. An example is a classifier output of 5% probability of GREEN, 65% probability of AMBER, and 30% probability of RED for a particular patient. Determining the optimal course of action in the face of such probabilities (e.g., treating the patient *as if* he were an AMBER) are the objectives of *decisional algorithms*, which directly utilize classifier outputs. In general, the subsequent decision must be a function of the classifier probabilities. The general theory of decisional algorithms is covered in Appendix D.

# 4 Mathematical Foundations of Polynomial Neural Networks

In both estimation and classification modeling, the most common approach is to express the output variable, $y$, assumed here to be a scalar, as an explicit function of the column vector of input variables, $\underline{X}$, *viz.*, $y = f(\underline{X})$. The inputs are quantities that are known or can be measured readily, whereas the output is either inaccessible to direct measurement or represents a future outcome. This is the basic conceptual approach of regression and polynomial neural network (PNN) modeling. In

the former, the strategy is to assume a generic functional form for $f(\underline{X})$, with certain parametric degrees of freedom (i.e., coefficients), and then to deduce values of each parameter in such a way that the model best emulates the "big picture" portrayed by a comprehensive database of historical cases *exemplifying* the actual relationship between $y$ and $\underline{X}$.

PNN methods are geared to precisely the same types of estimation and classification modeling problems to which regression methods have traditionally been applied. They espouse the same basic paradigm of relating outputs and inputs by way of an explicit function, *viz.*, $y = f(\underline{X})$. PNNs, however, offer a far more powerful and practical modeling methodology in that they largely overcome the greatest drawbacks in conventional regression approaches, as exemplified in Appendix G:

- Nonlinear regression requires that the model structure, i.e., construction of the synthetic input vector, $\underline{x}$, be stipulated *a priori*. In practical modeling problems, however, the "appropriate" structure is almost never known beforehand. The variety of potential model structures is so vast that a systematic trial-and-error search of alternative structures, i.e., stepwise regression, quickly proves prohibitively lengthy in many real-world scenarios involving large numbers of raw input variables.
- Such limitations usually restrict the analyst to simple functional forms, i.e., low-order polynomials, as candidate model structures, even though they may not necessarily describe the phenomenon accurately or to a sufficient degree of precision.
- Numerical determination of coefficient values in both least-squares and logistic regression requires inversion of a $P \times P$ matrix, where $P$ is the size of the synthetic input vector. This is a computationally intensive and cumbersome operation. More seriously, $P$ grows so rapidly as ever more complex polynomial structures are considered that it may even exceed the size of the training database, in which case the least-squares system of algebraic equations is underconstrained, rather than overconstrained. In other words, proper testing of such complex models requires additional training data.
- Comparison of alternative regression model structures requires cross-validation analysis, as demonstrated in Appendix G with random partitions of the database. Such tests must be performed repetitively to obtain good statistical assessments of model robustness. This, however, only exacerbates the computational difficulties, since it requires many more matrix inversions and work with even smaller training databases.

## 4.1 Group Method of Data Handling

The first brilliant stroke of insight into how such limitations could be overcome was from cyberneticist Alexei Ivakhnenko, whose *group method of data handling* (GMDH) methodology is the at the heart of PNN algorithms. GMDH is the key intellectual innovation that overcomes most of the major drawbacks of regression methods and makes PNNs fundamentally different and superior. The principal virtues of GMDH, in brief, are that it: (1) abandons any preconceived bias toward a particular model structure; (2) relieves the analyst from the burden of having to stipulate any such structures; (3) enables higher-order polynomials, and thus more general functional forms, to be considered; and (4) divides the mathematical labor of constructing such complex models over many nodes and layers.

The basic strategy of GMDH is as follows. First, quadratic polynomials are constructed from pairs of raw input variables. This is done for all such input pair combinations, and coefficients of the

quadratic forms are computed by regressing them against $y$. The estimates of $y$ produced by those quadratic polynomials form the first generation, or *layer*, of synthetic variables (called *nodes*). A cross-validation test determines which combinations are most useful and which are least useful; the least useful ones are discarded, or *carved away*. After the first layer has been constructed, a second layer of synthetic variables is formed by regressing quadratic forms in pairs of synthetic variables in the first layer against $y$. The process is repeated until the best synthetic variable in some layer gives a sufficiently accurate estimation of $y$. Note that the resulting *Ivakhnenko polynomial*, $I(\underline{X})$, which expresses $y$ in terms of the original inputs, $\underline{X}$, is of order $2^{H+1}$, in which $H$ is the number of *hidden layers* between the raw inputs and $y$.

Fig. 1 illustrates the generic architecture of a multilayer GMDH structure. The particular structure shown consists of four inputs, two hidden layers, and a single output from each node. This structure closely resembles feedforward *multilayer perceptron* (MLP) and other neural network

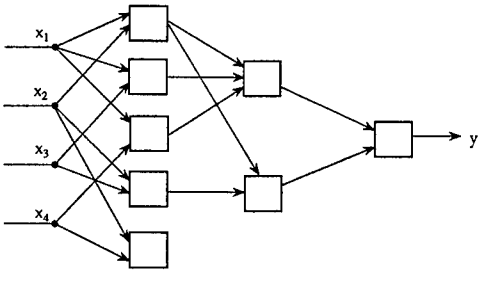


**Figure 1: Generic Architecture of GMDH Structure**

modeling structures. GMDH is the preeminent example of such a *connectionist* approach to multivariate data modeling. In the MLP, perceptrons, or *nodes*, correspond to Ivakhnenko's hidden-layer variables in that they each perform a relatively simple task, namely computing a quadratic function from just two inputs.

## 4.2 *GNOSIS*

*GNOSIS* (*Generalized Networks for Optimal Synthesis of Information Systems*) is a software package developed in house by BAI for synthesizing feedforward and recurrent neural networks. It was used in deriving all of the performance results presented herein. For estimation and classification problems, it incorporates the basic GMDH paradigm of model construction through hidden layers [44, 45], but with many important refinements over GMDH. These include post-transformation of nodal outputs, global optimization of coefficients in the Ivakhnenko polynomial, and relaxation of the rules for feeding estimates forward to the succeeding layer (e.g., combination of three or more inputs, usage of original inputs beyond the first layer, cubic nodal polynomials). *GNOSIS* can further refine layers by creating additional nodes in them whose inputs are not only outputs from the previous layer but also outputs from nodes within the current layer that have already been generated. This technique, known as *projection pursuit* [28], strengthens substantially the performance of the resulting PNN models. *GNOSIS*, moreover, harnesses the flexibility of polynomial basis functions to model *arbitrary* functions [31, 32]. The series expansion (and thus the nodal element) is sufficiently general to implement a variety of basis functions in addition to polynomials, including splines, wavelets, exponentials, and trigonometric functions.

*GNOSIS* obviates the need for cross-validation required by GMDH by appealing to the *predicted squared error* (PSE) criterion developed by Barron [2] for evaluating model structures. PSE is defined as

$$\mathrm{PSE} = \mathrm{FSE} + \frac{2K}{N\sigma^2} \tag{1}$$

in which FSE is the *fitted squared error* (i.e., the cumulative squared estimation error), $K$ is the number of degrees of freedom in the model (coefficients or nodes), $N$ is the number of exemplars in the training database, and $\sigma$ is a parameter whose appropriate value can be ascertained from statistical arguments, such as the Akaike Information Criterion (AIC). The second term on the right-hand side of Eq. 1 is a *complexity penalty* that describes the uncertainty in the coefficient values. From inspection of the full database, PSE appeals to the AIC to weigh the tradeoff between model performance (i.e., FSE) and complexity penalty. The model structure with least PSE generally corresponds to that which would be discovered via the much more arduous process of cross-validation.

With its accumulated arsenal of techniques, *GNOSIS* has a long history of successful industrial applications and is akin to many single-layer composition-of-functions techniques that are becoming increasingly popular among statisticians, such as *multivariate adaptive regression splines* (MARS) [3, 29].

*GNOSIS* is the outgrowth of earlier BAI software products (e.g., *ASPN-IIc* for estimation, *CLASS* for classification) and represents the accumulation of four decades of accumulated experience in the application of PNN methods in the commercial and industrial realms, both at BAI and its forerunner, Adaptronics, Inc.

*GNOSIS* overcomes the major weaknesses of regression and classical GMDH modeling and thus offers valuable benefits to the user. Most importantly, it judges the appropriate level of model complexity through internal criteria, while the user need only stipulate the output variables and *potential* input variables (some may be carved away entirely if they prove irrelevant). *GNOSIS* automatically infers the best network structure, node types, and coefficients from the data. The model is grown, through hidden layers, from the simplest possible form to a level of just-sufficient complexity, in view of functional relationships in the data and the quantity of data. This is preferable to postulating *a priori* model structures, which tend to have excessive degrees of freedom, and overfit on training data. Finding the best structural form is important to establish parsimonious models and to ensure thereby that the model performs well to new data not encountered during training. *GNOSIS* offers distinctive advantages in speed, degree of analyst involvement, and model accuracy that may be summarized as follows:

- *GNOSIS* offers a fast synthesis algorithm by spreading the mathematical labor over many nodes. It builds models on an element-by-element basis using nodal building blocks that are quadratic or cubic polynomials with no more than three inputs. In fact, the nodal basis functions can assume nonpolynomial forms such as wavelets, splines, or exponentials (in which case the "polynomial" in PNN is a misnomer). Construction of the model in this way reduces the likelihood of premature cessation of the optimization process.
- The internal PSE criterion governs the carving process and determines when to curtail network growth. This information-theoretic criterion reduces greatly the need for cross-validation, thus enhancing synthesis speed and, in principle, allows the entire database to be used for training. Need for analyst involvement is greatly reduced.

- The model structure search process incorporated by *GNOSIS* is extremely efficient and effective. It optimizes the structure of each layer before proceeding to create subsequent layers. This has the advantages that: (1) model performance upon completion of each layer is examined before moving on to a more complex structure; and (2) the more complex candidate structures accruing to an extra layer do not have to be generated from scratch. Only a few parameters are fitted at any given step in the process. In stepwise regression, every candidate structure must be refitted from scratch unless certain coefficient values are "frozen" in certain steps. This severely limits the space of potential models that are traversed during the regression and risks selection of models that may, in fact, not be very good at all.
- *GNOSIS* utilizes projection pursuit and outputs prior to the previous layer at any step as candidate nodal inputs. In this way, complex interrelationships can be discovered using relatively simple functions, and superior models generally emerge.
- *GNOSIS* uses a Gauss-Newton technique regularized using the Levenberg-Marquardt algorithm to learn the coefficients of arbitrary linear and nonlinear models that optimize network performance with respect to arbitrary cost functions. Thus, *estimation* and *classification* networks can both be synthesized using the same tool.

# 5 Analysis of Historical UVA Pre-Hospital Data

The present section marks the beginning of our discussion and analysis of trauma registry data to which we obtained access in Phase I.

## 5.1 University of Virginia Hospital Trauma Data

As the core of BAI's Phase I research efforts, we demonstrate herein a rigorous process for deriving and testing mortality prediction models. PNN models were synthesized using *GNOSIS* for both estimation (minimum squared error) and classification (minimum logistic error) objectives. For purposes of testing algorithmic methods on such data, we acquired access to the internal trauma registry of the University of Virginia Hospital, covering the 30-month period from January 1, 1994 to June 30, 1996. This database, as provided to BAI by Dr. George Lindbeck, included the (uncoded) standard criteria of AGE, SEX, B/P, EY, VB, MT, RR, SBP, ISS (with AIS breakdown), and survival outcome. The database contained a total of 4,436 patients received by the Emergency Department at the Hospital. Confidential data fields were removed by the Hospital for purposes of this study. Among the records, several hundred were incomplete and were therefore discarded. We also excluded babies ($<$ 2 years of age); this left a total of 3,628 complete patient exemplars as the definitive body of empirical data we proceeded to analyze. From here on, we shall refer to this data set as the UVA database.

## 5.2 Distributions within the UVA Database

As the first step in inspecting the UVA database (after the preparatory steps above), we examined histogram plots of the distributions for the various fields. As a dichotomous variable, the gender distribution was 1,338 females (36.9%) and 2,290 males (63.1%). Among all patients, 407 had penetrating injuries (11.2%) and 3,183 had sustained blunt trauma (88.8%). 3,486 patients

survived (96.1%) and 142 died (3.9%). All of the other variables, which were nondichotomous, assumed a wide range of values, histogram plots for which are provided in Fig. 2.

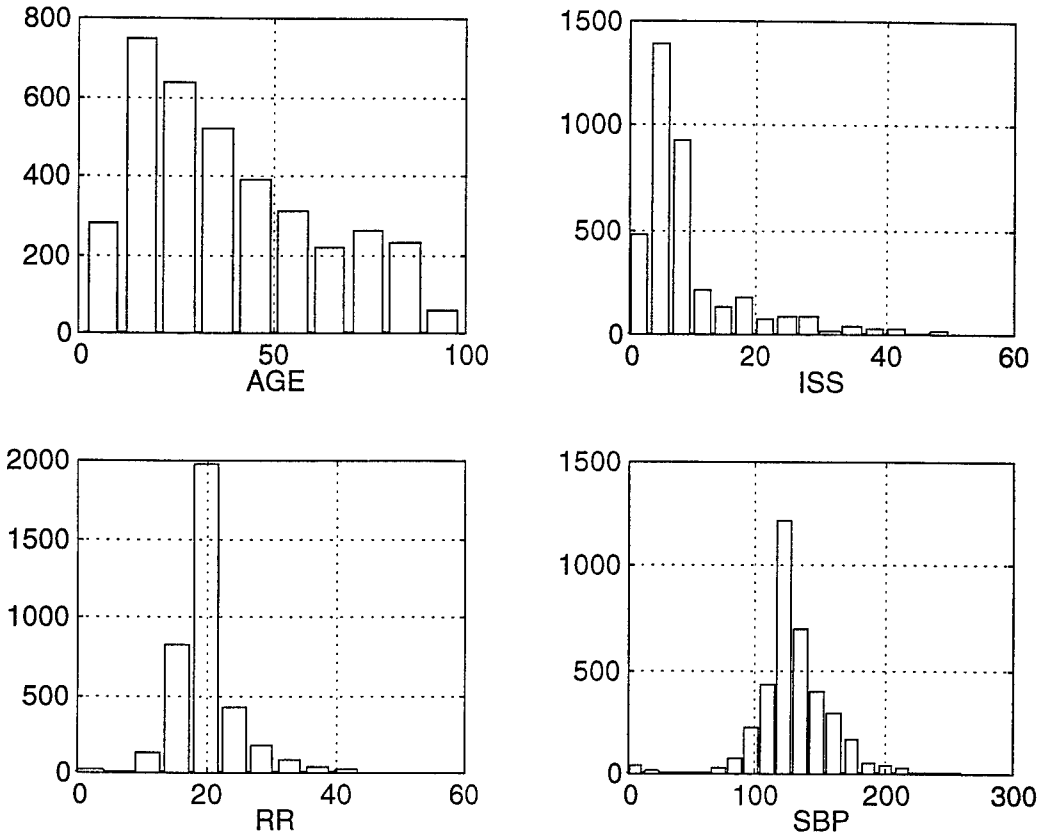


**Figure 2: Univariate Distributions for the UVA Database**

The age distribution is comprehensive and demographically representative, with a preponderance of patients in the teens, twenties, and thirties. The distribution is slightly bimodal in that it flattens in the elderly group. Intuitively, this pattern makes sense, with large numbers of young adult trauma cases due to automotive and athletic accidents and a disproportionate number of elderly cases due to falls, hip fractures, and the like.

The Injury Severity Score (ISS) distribution appears unimodal and Poisson-like, with a very large concentration of mildly and moderately injured patients with scores between 5 and 10. Although not shown, the highest ISS was 66, which is not far below the maximum possible score of 75. Respiratory rate (RR) and systolic blood pressure (SBP) are unimodal and, to a crude approximation, normally distributed, with means and standard deviations of $19.71 \pm 5.66$ breaths per minute and $131.30 \pm 26.69$ millimeters of mercury (mm Hg) respectively. The chief anomalies in the RR and SBP distributions are an appreciable skewness in RR and distinct groups of patients in which blood pressure or respiration had completely vanished; large majorities of patients in both

of these group died.

The neurological variables EY, VB, and MT require special consideration because their distributions are extremely skewed, with a large majority of patients (3,204 total, or 88.3%) having a total Glasgow Coma Scale of GCS = 15. Since GCS = 15 implies EY = 4, VB = 5, and MT = 6 for each patient in this group, it follows that EY, VB, and MT are also strongly skewed. The set of other patients was divided approximately evenly between slight impairment (GCS = 14), total unconsciousness (GCS = 3), and the intermediate zone ($4 \leq \text{GCS} \leq 13$). The numbers of incidents respectively in these groups were 147 (4.1%), 132 (3.6%), and 145 (4.0%). Uneven distributions such as this must be recognized in fitting models to the database and may sometimes present difficulties, such as in attempting to test such fitted models on a population with a much larger percentage of low GCS scores. For GCS, this point is particularly important because it appears to be the single most effective indicator of both injury severity and survival chances.

## 5.3 Univariate Analyses

Having examined the distributions of variables in the database, we look next for patterns of correlation between the independent input variables, namely SEX, AGE, B/P, EY, VB, MT, RR, and SBP, and the dependent output variables, namely ISS and survival outcome. ISS is treated as an output variable because even though it reflects anatomical damage incurred at the time of injury, the extent of damage that it reflects cannot be ascertained on the spot, at least not with current biomedical technology. To the contrary, ISS can only be ascertained retrospectively based on the findings of emergency department surgeons. Survival outcome is treated as a dependent variable for much more obvious reasons.

Table 1 provides the mortality (percentage of patients in a given subset who died) and the mean ISS for groups of patients such that the tabulated input variable lies in the indicated range. Results are partitioned into bins and provided for each of the six input variables. The first row in the SEX table, for instance, indicates that 2.91% of the 1,338 female patients in the database died and that the average ISS in that group was 8.66. Such univariate analyses serve the purpose of identifying patterns in the database. Such patterns may, in certain cases, reflect causal connections between variables; in other cases, they may reflect merely indirect or coincidental correlations that have no underlying significance. A modicum of caution, therefore, must be exercised in the interpretation of the tabulated statistics.

**Table 1: Univariate Analysis of Input Variables**

**SEX**

| Range | Mortality | ISS |
|---|---|---|
| F | 0.0291 | 8.66 |
| M | 0.0450 | 8.84 |

**B/P**

| | | |
|---|---|---|
| B | 0.0373 | 8.93 |
| P | 0.0541 | 7.56 |

**AGE**

| | | |
|---|---|---|
| $< 10$ | 0.0097 | 6.35 |
| $10 - 19$ | 0.0330 | 8.80 |
| $20 - 29$ | 0.0441 | 9.02 |
| $30 - 39$ | 0.0238 | 8.74 |
| $40 - 49$ | 0.0219 | 8.58 |
| $50 - 59$ | 0.0327 | 8.09 |
| $60 - 69$ | 0.0488 | 9.18 |
| $70 - 79$ | 0.0695 | 9.68 |
| $\geq 80$ | 0.0906 | 9.79 |

**GCS**

| | | |
|---|---|---|
| 15 | 0.0091 | 7.22 |
| 14 | 0.0544 | 12.59 |
| $4 - 13$ | 0.2045 | 22.29 |
| 3 | 0.5379 | 26.84 |

**RR**

| Range | Mortality | ISS |
|---|---|---|
| $RR = 0$ | 0.8750 | 25.06 |
| $0 < RR < 10$ | 0.4000 | 27.20 |
| $10 \leq RR < 15$ | 0.0206 | 7.75 |
| $15 \leq RR < 20$ | 0.0172 | 6.88 |
| $20 \leq RR < 25$ | 0.0491 | 9.75 |
| $25 \leq RR < 30$ | 0.0729 | 12.64 |
| $RR \geq 30$ | 0.0788 | 15.18 |

**SBP**

| | | |
|---|---|---|
| $SBP = 0$ | 1.0000 | 25.44 |
| $0 < SBP < 80$ | 0.2821 | 20.51 |
| $80 \leq SBP < 90$ | 0.3226 | 21.06 |
| $90 \leq SBP < 100$ | 0.0792 | 11.70 |
| $100 \leq SBP < 110$ | 0.0351 | 9.96 |
| $110 \leq SBP < 120$ | 0.0221 | 8.11 |
| $120 \leq SBP < 130$ | 0.0147 | 7.21 |
| $130 \leq SBP < 140$ | 0.0190 | 8.01 |
| $140 \leq SBP < 150$ | 0.0220 | 8.22 |
| $150 \leq SBP < 160$ | 0.0291 | 9.25 |
| $160 \leq SBP < 170$ | 0.0469 | 9.03 |
| $170 \leq SBP < 180$ | 0.0476 | 10.01 |
| $SBP \geq 180$ | 0.0625 | 10.26 |

From the numbers in Table 1, the following comments may be made:

- Mortality and ISS with respect to AGE both exhibit a common pattern of bimodality. Young children had significantly better outcomes than adults. Mortality and ISS both rise in the twenties, decline in middle age, and peak in the elderly age groups. The bimodality is even more pronounced among the group of patients who died, as shown in Fig. 3.

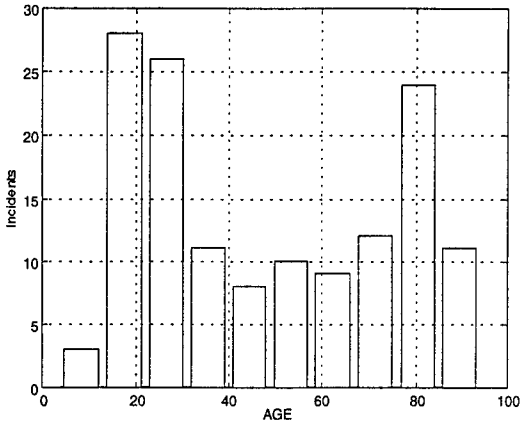


**Figure 3: Age Distribution among Nonsurvivors**

This is chiefly an *epidemiological* phenomenon in that people in their twenties and the elderly suffer disproportionate incidence rates of trauma; it does *not* imply that other things equal, individuals in these groups are more likely to die from injuries than, for example, people in their forties or fifties.

- That males exhibit a mortality rate more than 50% greater than females similarly does not imply that men are more vulnerable to trauma than women. Rather, it reflects the significant difference in the age distributions of the sexes, with the women (mean age 45.7) generally being older than the men (mean age 35.4), and thus, the different causes of injury one would expect to find in the young and elderly groups.
- The results for B/P, surprisingly, are mixed. Whereas patients with penetrating wounds had greater chance of dying than blunt trauma victims, their mean ISS was lower. The significance of B/P, from this initial impression, is thus inconclusive.
- GCS correlates very strongly with both mortality and ISS. As neurological function falls, mortality and ISS both rise sharply. Among the inputs, the Glasgow components are collectively the strongest indicators of outcome, and are therefore indispensible input variables in all candidate model structures.
- RR also exhibits a strong relationship with both mortality and ISS. Hypoventilatory patients (RR $< 10$), whose impairment of neurological function causes respiratory rate to fall, and those with airway blockage, are clearly in greater danger than those with normal breathing. Apnea, or complete cessation of breathing, correlates with grave injuries and is almost always fatal, as the table shows. There is also a significant tendency for outcomes to become less favorable at high RR levels (RR $> 25$). These may correspond to hypovolemic cases, in which respiratory and heart rates both increase. At intermediate, normal values of RR, mortality and ISS are both below average and are relatively flat.
- Much the same can be said about SBP, sharp departures from the normal range of which correlate with both higher mortality and ISS. As with RR, below-normal deviations tend to be much more serious than above-normal anomalies. Outcomes are especially bleak when SBP falls below 90, which is propitiously chosen as a cutoff by many conventional trauma scores, such as RTS and TTR. As with RR, the outcome is almost always (with the UVA data, not

even "almost") fatal when the value falls all the way to zero.

- Some clinical indicators may have to be modified in order to compare demographically distinct groups. For example, there is a direct correlation between AGE and SBP; blood pressure generally increases with age. The relationship, based on linear regression of the UVA data, is SBP $\approx 118.5 + 0.339 \times$ AGE, as shown in Fig. 4. It may therefore be of some help to use an age-adjusted SBP, for example, instead of the raw data in the models. PNN synthesis, however, accomplishes this automatically.

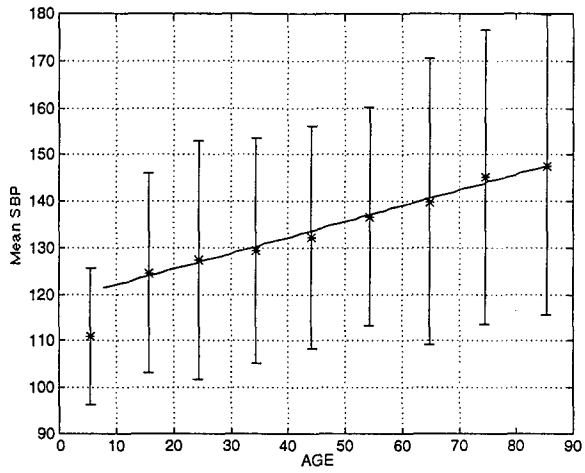


**Figure 4: SBP vs. AGE (Error Bar = 1 Standard Deviation)**

Equally importantly, the two outcomes, mortality and ISS, themselves are mutually correlated. The average ISS for survivors was $8.08 \pm 7.16$, whereas that for death cases was $25.80 \pm 12.82$. Despite the broad variances, the survivor and nonsurvivor classes are significantly different and distinguishable in ISS levels. The mortality as a function of ISS is tabulated in Table 2.

**Table 2: Mortality vs. ISS**

| Range | Mortality |
|---|---|
| ISS $<$ 10 | 0.0085 |
| 10 $\leq$ ISS $<$ 20 | 0.0318 |
| 20 $\leq$ ISS $<$ 30 | 0.2206 |
| 30 $\leq$ ISS $<$ 40 | 0.3485 |
| 40 $\leq$ ISS $<$ 50 | 0.3913 |
| ISS $\geq$ 50 | 0.7692 |

Clearly, mortality increases directly and very rapidly with ISS. This indicates that if ISS could be determined in the field, either directly through superior biomedical instrumentation or indirectly through use of an inference algorithm, one might have a very powerful indicator of survival

**Table 3: Survival Outcome Classification Models**

| Ages | Degree of Nodal Polynomials | Discrimination Power | Area under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|---|
| all | 1 | 0.887 | 0.939 | 1198.7 | 602.2 | 0.502 | 0.0001 |
| | 2 | 0.891 | 0.948 | 1198.7 | 555.4 | 0.403 | 0.0001 |
| | 3 | 0.908 | 0.958 | 1198.7 | 461.5 | 0.385 | 0.0002 |
| <55 | 1 | 0.918 | 0.947 | 720.6 | 282.9 | 0.393 | 0.0146 |
| | 2 | 0.944 | 0.972 | 720.6 | 247.2 | 0.343 | 0.0089 |
| | 3 | 0.955 | 0.964 | 720.6 | 195.0 | 0.271 | 0.0253 |
| ≥55 | 1 | 0.806 | 0.872 | 455.4 | 297.3 | 0.653 | 0.0072 |
| | 2 | 0.823 | 0.886 | 455.4 | 260.8 | 0.573 | 0.3129 |
| | 3 | 0.847 | 0.934 | 455.4 | 191.4 | 0.420 | 0.2103 |

prospects. In the following subsection, we pursue development of such algorithmic capabilities.

## 5.4 PNN Classification Models for Survival Outcome

As the first modeling effort, we constructed classifiers (nonlinear logistic regression models) to learn survival outcome as a function of seven inputs (AGE, B/P, EY, VB, MT, RR, and SBP) using *GNOSIS*. PNN models were trained on: (1) the full database of 3,628 exemplars; (2) on the subset of patients under age 55 years (2,699 exemplars); and (3) on patients of age 55 or older (929 exemplars). This was done to segregate the two dissimilar age groups in the nonsurvivor group. Age 55 was chosen as the dividing line between the two age groups because it adequately separates the two groups in Fig. 3 and also matches the cutoff that TRISS (Appendix A.15) uses in coding the AGE variable. This facilitates subsequent analysis of TRISS performance on the UVA data. For each of the three age-group sets, *GNOSIS* models with nodal polynomials of first, second, and third degree were synthesized. Salient performance statistics are tabulated in Table 3. The output of the *GNOSIS* classifiers are the probability, $P_d$, that the patient will die. By varying the threshold placement, $\xi$, vis-à-vis that probability, i.e., such that $0 < \xi < 1$, a *receiver-operating characteristic* (ROC) curve (Appendix D.2) is obtained. The ROC curve for a given classification model shows the specificity-sensitivity characteristics that can be achieved with various threshold settings for decisional algorithms. A key summary index for the discrimination capability implied by the curve is that level of specificity at which equal sensitivity can be obtained. We shall refer to this common value of specificity and sensitivity as the *discrimination power* of a dichotomous decisional algorithm. The first row in Table 3, for instance, indicates that a *GNOSIS* classifier model with linear nodal polynomials, trained and evaluated on the full database encompassing all age groups, can achieve simultaneous specificity and sensitivity of 88.7%. The area under the curve, which varies from 0.5 for complete indistinction (i.e., a 45° line) to 1 for perfect discrimination, also provides a good, but somewhat less reliable, indicator of ROC curve quality. Unlike discrimination power, which focuses on just one threshold setting, the area statistic describes the ROC curve *in toto*. Fig. 5 illustrates a ROC curve plot for the third-degree *GNOSIS* classification model (all ages) in Table 3. We shall refer to this model, which represents our main result for pre-hospital triage algorithms, as Model I.

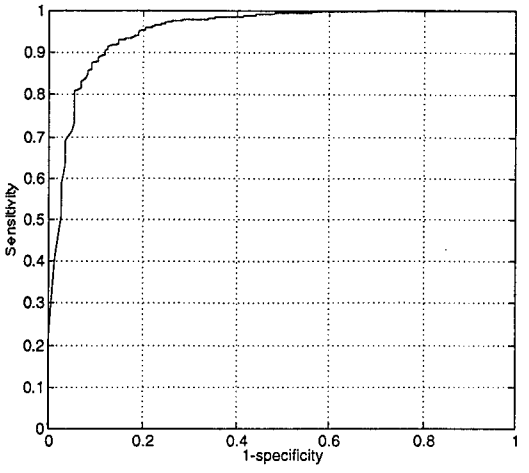


**Figure 5: ROC Curve for Model I**

The results in Table 3 provide summary statistics of how well *GNOSIS* predicts outcome for the UVA data. The younger age group evidently lends itself to much easier mortality prediction than does the older age group, with discrimination powers of 95.5% and 84.7% respectively. The results also indicate that performance can be improved by use of higher-degree nodal polynomials in the *GNOSIS* synthesis. The 91% discrimination power among both age groups combined is comparable to the best specificity-sensitivity results reported in the literature for any conventional scoring system, such as TTR [8].

Table 3 provides some additional classification performance statistics. The log-likelihood, $\Lambda$, provides an indicator of classification model performance. For a logistic regression or PNN classification model, the log-likelihood is defined as

$$\Lambda = \sum_{i=1}^{N} \sum_{c=1}^{C} (y_i = c) \ln \pi_{i,c} \tag{2a}$$

in which $\pi_{i,c}$ is the probability, according to the model, that the $i$'th observation belongs to class $c$. This figure of merit should be compared to the baseline log-likelihood, $\Lambda_0$, obtained by using the *a priori* probabilities, $\alpha_c$, in place of $\pi_{i,c}$, *viz.*

$$\Lambda_0 = \sum_{i=1}^{N} \sum_{c=1}^{C} (y_i = c) \ln \alpha_c = N \sum_{c=1}^{C} \alpha_c \ln \alpha_c \tag{2b}$$

It follows from the rightmost expression in Eq. 2b that $\Lambda_0$ is a measure of the entropy of the class distribution in the training database. The ratio $\Lambda/\Lambda_0$ provides a measure of the extent to which the classification model "explains" the tendency of various exemplars to belong to different classes. It is analogous to the $R^2$ statistic in least-squares regression and PNN counterparts for estimation modeling, wherein $\Lambda$ is analogous to the sum of the squared estimation errors, $\sum_i (y_i - \hat{y}_i)^2$, and $\Lambda_0$ is analogous to the sum of the squared deviations from the mean, $\sum_i (y_i - \overline{y})^2$. A small $\Lambda/\Lambda_0$ ratio indicates superior classification, which accrues to both intrinsic separability of the classes and the degree to which the classification model effectively exploits the input variables provided to it.

Finally, the quantity $p$ tabulated in the right-hand column is the Hosmer-Lemeshow [70] probability statistic for *goodness-of-fit*, which indicates the extent to which the class membership probabilities (in dichotomous problems) account accurately for the actual outcomes encountered in the

**Table 4: Cross-Validation of Model I**

| Degree of Nodal Polynomials | Discrimination Power under Self-Validation | Discrimination Power under Cross-Validation |
|---|---|---|
| 1 | 0.887 | 0.892±0.006 |
| 2 | 0.891 | 0.898±0.007 |
| 3 | 0.908 | 0.882±0.014 |

training database, especially in "gray" regions where the probabilities are close to neither zero nor unity. It partitions the data into bins (customarily ten) based on the model-generated probabilities (usually via logistic regression) and compares the observed distribution of exemplars over the bins with the expected numbers of exemplars falling into each bin, or *group*. In the case of ten groups, for example, the bins are deciles of model-computed probability between 0 and 1. For each group, the *expected* number of death incidents is $N_i\overline{\pi}_i$, in which $N_i$ is the number of exemplars in the $i$'th group and $\overline{\pi}_i$ is the mean predicted probability of death for exemplars in that group. The expected number of deaths is compared to the *observed* number, $O_i$, of deaths in the $i$'th group. The Hosmer-Lemeshow goodness-of-fit statistic is then computed as:

$$\chi_{\text{HL}} = \sum_{i=1}^{g} \frac{(O_i - N_i\overline{\pi}_i)^2}{N_i\overline{\pi}_i(1-\overline{\pi}_i)} \tag{3}$$

in which $g$ is the number of groups. The goodness-of-fit statistic is compared to a $\chi^2$-distribution with $g-2$ degrees of freedom, and a null-hypothesis probability, $p$, that the null model ($\pi_d = \alpha_d$, in which $\alpha_d$ is the percentage of actually deaths in the training database) describes the data as well as the proposed model. A small $p$ value indicates better explanatory power than a large value. $p$ is not computed if no more than two groups can feasibly be created. The statistic for Model I is extremely good this respect.

## 5.5 Validation of PNN Classification Models

As in stepwise regression, model validation customarily entails cross-validation using repetitive random partitions of the database. To double-check the Model I results via cross-validation, we performed 30 random 90%-10% training-evaluation partitionings of the UVA database and tabulated the mean and standard error thereof. This was done for the first-, second-, and third-degree models on all ages. The results are tabulated in Table 4. The discrimination power statistics in Table 3 were for self-validation, in which the model was trained on the entire database and then evaluated on that *same* data. Although this practice is widely recognized as an anathema, for reasons explained in the discussion of stepwise regression in Appendix G, self-validation is actually not unreasonable in the context of *GNOSIS* modeling since, as indicated previously, the PSE criterion automatically prevents overfitting. Thus, one expects models synthesized using *GNOSIS* to perform equally well on test data sets having exemplar distributions similar to those of the training data set, assuming that good training database construction techniques are follows (see, e.g., [5]). Under such practices, the PNN models are "interpolating" rather than "extrapolating."

That this is, in fact, the case is demonstrated in Table 4, in which the self- and cross-validation results are compared. The means of the cross-validation distributions are nearly equal to the

discrimination powers obtained from self-validation. The results for the first-, second-, and third-degree models indicate that the self-validated results are statistically within the margin of error of the cross-validated results. The cross-validated mean discrimination powers for the first two nodal polynomial types actually exceed their self-validated counterparts.

Moreover, the uncertainty in the cross-validated means, as represented by the standard deviations make use of these figures somewhat unreliable for comparing different classification models unless many random partition cuts are done. The self-validated performance statistics, however, are stable and serve as good benchmarks for comparing model performance. Even if, for example, the self-validated result does tend to overstate the performance results that would be found from exhaustive cross-validation, that overstatement will at least be consistent. The self-validated results, owing to PSE, furnish an efficient means of summarily comparing the performance results of different models.

## 5.6 Classification Performance of Conventional Scores on UVA Data

We next assess the classification performance of several conventional pre-hospital scores as a baseline. For this purpose, we selected a certain subset of the scores (namely TTR, KRC, TS, T-RTS, CRAMS, and RSM) documented in Appendix A. Others could not be computed since they involve mechanistic (such as in MOI) or physiological criteria (heart rate in PHI and RTI) that were not provided in the UVA database records. Even among some of the scores that were selected, however, there are certain criteria (e.g., respiratory effort, tenderness of abdomen) that cannot be deduced automatically from inspection of the UVA data. Only T-RTS and RSM could readily be computed, but for the other four, it was necessary to make certain reasonable assumptions as to filling in fields that could not be ascertained readily. The following such assumptions were made in computing the scores:

- **TTR**
  B/P=1 and ($AIS_1 \geq 1$ or $AIS_3 \geq 1$) for "penetrating cranial, neck, or thoracic injury"
- **KRC**
  Eye opening criterion interpreted as $EY \leq 3$
  $SBP \leq 90$ used for abnormal capillary refill
  B/P=1 and ($AIS_1 \geq 1$ or $AIS_3 \geq 1$) for same anatomical criterion as in TTR
- **TS**
  Respiratory effort assumed normal if $10 \leq RR \leq 24$, shallow/retractive otherwise
  Capillary refill = 2 if $SBP \geq 90$, 1 if $0 < SBP < 90$, 0 if $SBP = 0$
- **CRAMS**
  Capillary refill disregarded in circulation criterion
  Respiratory effort normal if $10 \leq RR \leq 24$, none if $RR = 0$, shallow/retractive otherwise
  Abdomen = 2 if $AIS_3 + AIS_4 = 0$, 1 if $AIS_3 + AIS_4 = 1$, 0 otherwise
  Motor = 2 if $MT = 6$, 1 if $2 \leq MT \leq 5$, 0 if $MT = 1$
  Speech = 2 if $VB = 5$, 1 if $3 \leq VB \leq 4$, 0 if $1 \leq VB \leq 2$

Tables 5, 6, and 7 give the specificity and sensitivity as functions of threshold placement for each of the six selected conventional scores, when tested on the UVA data. Since these scoring systems do not involve class membership probabilities, the log-likelihood and the Hosmer-Lemeshow statistics are not applicable. The scores themselves serve as thresholds, e.g., a TTR score of 1 indicates a decisional output, not a probability.

It is evident from the results in Tables 5-7 that none of the six conventional scores that we computed perform satisfactorily on the UVA data. They all fail to achieve good specificity-sensitivity characteristics on the full database and perform extremely poorly on the older age group. Sensitivities are especially bad. The only "good" results, with specificity and sensitivity both above 90%, are achieved by TTR, TS, T-RTS, and CRAMS in the younger age group, but all of these underperform the 95.5% discrimination power accruing to the *GNOSIS* classification model described above.

**Table 5: Performance of TTR (left) and KRC (right)**

| Ages | Specificity | Sensitivity |
|---|---|---|
| all | 0.9509 | 0.7113 |
| $< 55$ | 0.9416 | 0.9000 |
| $\geq 55$ | 0.9792 | 0.4677 |

| Ages | Specificity | Sensitivity |
|---|---|---|
| all | 0.9730 | 0.4085 |
| $< 55$ | 0.9683 | 0.5000 |
| $\geq 55$ | 0.9873 | 0.2093 |

**Table 6: Performance of TS (left) and T-RTS (right)**

| Ages | Triage Rule | Specificity | Sensitivity |
|---|---|---|---|
| all | TS $\leq 15$ | 0.8551 | 0.8239 |
| | TS $\leq 14$ | 0.8649 | 0.8099 |
| | TS $\leq 13$ | 0.9495 | 0.7465 |
| $< 55$ | TS $\leq 15$ | 0.8450 | 0.9375 |
| | TS $\leq 14$ | 0.8557 | 0.9375 |
| | TS $\leq 13$ | 0.9420 | 0.9125 |
| $\geq 55$ | TS $\leq 15$ | 0.8858 | 0.6774 |
| | TS $\leq 14$ | 0.8927 | 0.6452 |
| | TS $\leq 13$ | 0.9723 | 0.5323 |

| Ages | Triage Rule | Specificity | Sensitivity |
|---|---|---|---|
| all | T-RTS $\leq 11$ | 0.9068 | 0.7958 |
| | T-RTS $\leq 10$ | 0.9610 | 0.7394 |
| | T-RTS $\leq 9$ | 0.9727 | 0.6549 |
| $< 55$ | T-RTS $\leq 11$ | 0.8961 | 0.9375 |
| | T-RTS $\leq 10$ | 0.9546 | 0.9125 |
| | T-RTS $\leq 9$ | 0.9679 | 0.8500 |
| $\geq 55$ | T-RTS $\leq 11$ | 0.9389 | 0.6129 |
| | T-RTS $\leq 10$ | 0.9804 | 0.5161 |
| | T-RTS $\leq 9$ | 0.9873 | 0.4032 |

**Table 7: Performance of CRAMS (left) and RSM (right)**

| Ages | Triage Rule | Specificity | Sensitivity |
|---|---|---|---|
| all | CRAMS $\leq 9$ | 0.6506 | 0.9083 |
| | CRAMS $\leq 8$ | 0.7493 | 0.8451 |
| | CRAMS $\leq 7$ | 0.9091 | 0.8028 |
| $< 55$ | CRAMS $\leq 9$ | 0.6518 | 0.9750 |
| | CRAMS $\leq 8$ | 0.7430 | 0.9625 |
| | CRAMS $\leq 7$ | 0.9000 | 0.9500 |
| $\geq 55$ | CRAMS $\leq 9$ | 0.6471 | 0.8226 |
| | CRAMS $\leq 8$ | 0.7682 | 0.6935 |
| | CRAMS $\leq 7$ | 0.9366 | 0.6129 |

| Ages | Triage Rule | Specificity | Sensitivity |
|---|---|---|---|
| all | RSM $\leq 11$ | 0.8617 | 0.8169 |
| | RSM $\leq 10$ | 0.9504 | 0.7113 |
| | RSM $\leq 9$ | 0.9713 | 0.6549 |
| $< 55$ | RSM $\leq 11$ | 0.8519 | 0.9375 |
| | RSM $\leq 10$ | 0.9427 | 0.8875 |
| | RSM $\leq 9$ | 0.9668 | 0.8500 |
| $\geq 55$ | RSM $\leq 11$ | 0.8916 | 0.6613 |
| | RSM $\leq 10$ | 0.9735 | 0.4839 |
| | RSM $\leq 9$ | 0.9850 | 0.4032 |

## 5.7 Comparison of RTS Coefficients derived from UVA Database to Champion-Sacco Values

Beyond the conventional scores just evaluated, we also explored PNN methods vis-à-vis two well-established algorithms, namely TRISS and ASCOT, for retrospective determination of attributes distinguishing patients who generally survive and those who tend not to survive. TRISS computes a probability of survival for a given patient based on a logistic formula in RTS, ISS, and age ($\mathrm{AGE}_c = 1$ if the patient is age 55 or older, zero otherwise). Since ISS cannot be ascertained before hospitalization, TRISS cannot be employed, as such, in the pre-hospital environment. However, it suggests that it might be possible to develop an on-line TRISS-like algorithm to predict survival outcome in terms of an *estimated* ISS value and the standard pre-hospital inputs. As the first step in developing such an algorithm, however, we must first treat the RTS component, which is a linear combination of coded values of SBP, RR, and GCS. The coefficients in the established definition of RTS, however, are based on a logistic regression that Champion and Sacco [20] performed on the American MTOS database. To construct a TRISS-like model for pre-hospital use, it therefore makes sense to obtain new RTS coefficients by performing the same logistic regression on the UVA data, in which a linear polynomial of the form

$$\underline{\theta}^{\mathrm{T}} \cdot \underline{x} = \theta_0 + \theta_{\mathrm{GCS}} \times \mathrm{GCS}_c + \theta_{\mathrm{SBP}} \times \mathrm{SBP}_c + \theta_{\mathrm{RR}} \times \mathrm{RR}_c \tag{3a}$$

is sought, in which

$$P_s = \frac{1}{1 + e^{-\underline{\theta}^{\mathrm{T}} \cdot \underline{x}}} \tag{3b}$$

is the probability of survival. The resulting coefficients and the corresponding Champion-Sacco values are tabulated in Table 8 The computed coefficients are in fair agreement with the corresponding

**Table 8: MTOS- and UVA-derived RTS Coefficient Values**

| Term | MTOS Values | UVA Values (all ages) | UVA Values ($< 55$) | UVA Values ($\geq 55$) |
|---|---|---|---|---|
| $\theta_0$ | $-3.5718$ | $-5.0095$ | $-3.7883$ | $-11.4422$ |
| $\theta_{\mathrm{GCS}}$ | 0.9368 | 1.0587 | 1.3104 | 1.1301 |
| $\theta_{\mathrm{SBP}}$ | 0.7326 | 0.9667 | 0.9207 | 1.0843 |
| $\theta_{\mathrm{RR}}$ | 0.2908 | 0.3311 | 0.1174 | 1.4989 |

Champion-Sacco values, except for the constant and RR terms in the older group. Statistics for the resulting ROC curves are given in Table 9. The two rows in each age group give the performance results on the UVA data with the Champion-Sacco and UVA-derived RTS coefficients in the survival outcome model. The results indicate that the UVA-fitted coefficients do not yield significantly better performance than the Champion-Sacco values. As was the case above, the models perform better on the AGE $< 55$ group but worse on the AGE $\geq 55$ group than on the general population. Although RTS appears to perform very well on the younger age group, it performs extremely poorly on the older group and fails to match any of the corresponding discrimination power benchmarks achieved by the *GNOSIS* model. The results are consonant with the findings in the previous section for the T-RTS triage rules. Where the Hosmer-Lemeshow $p$ value is not given, *SAS* was not able to form the needed data bins based on the model-generated probabilities.

**Table 9: ROC Curve Statistics for RTS Classifiers**

| Ages | Coefficient Set | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|---|
| all | MTOS | 0.8162 | 0.877 | 1198.7 | 712.4 | 0.594 | 0.0376 |
| | UVA | 0.8162 | 0.877 | 1198.7 | 687.6 | 0.574 | 0.0375 |
| $< 55$ | MTOS | 0.9368 | 0.954 | 720.6 | 329.9 | 0.458 | 0.1098 |
| | UVA | 0.9371 | 0.954 | 720.6 | 299.0 | 0.415 | 0.1301 |
| $\geq 55$ | MTOS | 0.7081 | 0.788 | 455.4 | 365.1 | 0.802 | – |
| | UVA | 0.7081 | 0.787 | 455.4 | 315.1 | 0.692 | – |

## 5.8 PNN Estimation Models for ISS

As the second step in constructing TRISS-like algorithms for pre-hospital use, we used *GNOSIS* to construct estimation models of ISS as a function of the clinical inputs AGE, BP, EY, VB, MT, RR, and SBP. This is the exact same set of inputs used in the *GNOSIS* classification model. It is noteworthy that, unlike in any of the conventional scores, this set utilizes the three Glasgow subcomponents separately and does not use any coded values for the variables (except for B/P) that are naturally described by a continuum scale. Comparison of the layer-by-layer progress of the PNN synthesis process with the results for least-squares regression highlights the performance edge accruing to the exploitation of flexible functional forms and the ability of *GNOSIS* to *discover* appropriate model structures.

The ISS estimation results for the separate age groups are shown in Table 10. It is somewhat

**Table 10: *GNOSIS* ISS Estimation in Age-Segregated Groups**

| 3rd-degree Nodal Polynomials | | |
|---|---|---|
| Layer | AGE $< 55$ | AGE $\geq 55$ |
| 1 | 6.323 | 5.784 |
| 2 | 6.230 | 5.566 |
| 3 | 6.183 | 5.465 |
| 4 | 6.172 | 5.422 |

| 4th-degree Nodal Polynomials | | |
|---|---|---|
| Layer | AGE $< 55$ | AGE $\geq 55$ |
| 1 | 6.227 | 5.611 |
| 2 | 6.099 | 5.424 |
| 3 | 6.034 | 5.299 |
| 4 | 6.015 | 5.206 |

surprising, in view of the preceding results for survival prediction, that it is easier to infer ISS in the older age group. The RMS estimation error for ISS in the older group is smaller despite the tendency of that group to have slightly higher ISS scores. This may conceivably reflect that a greater diversity of injury types and severities within the younger age group.

## 5.9 ISS as a Decisional Threshold

As was observed in the preliminary inspection of the UVA database, ISS correlates strongly with survival outcome. Among hospitalized patients, ISS also correlates strongly with operational definitions of *major trauma* in terms of postoperative findings (AIS and ICD-9 scores) and the intensity of care required or attempted. In retrospective analyses, a summary definition of major

trauma as ISS $\geq$ 16 is often interpreted as those patients who "should have been" identified (in the field or in hospital) as critical. The direct usage of ISS as a decisional threshold in this manner suggests the possibility of a *GNOSIS*-generated ISS estimate serving as a pre-hospital algorithm. The performance results, using projection-pursuit estimation models with 4th-degree nodal polynomials, are given in Table 11. Performance results with the actual and PNN-estimated

**Table 11: ISS-based Decision Rules**

| | Actual ISS | | | Estimated ISS | | |
|---|---|---|---|---|---|---|
| Ages | Area under ROC Curve | Discrimination Power | Threshold Value | Area under ROC Curve | Discrimination Power | Threshold Value |
| all | 0.882 | 0.8209 | 13.36 | 0.882 | 0.8209 | 13.36 |
| $< 55$ | 0.947 | 0.8927 | 17.58 | 0.940 | 0.9125 | 12.98 |
| $\geq 55$ | 0.784 | 0.7198 | 9.78 | 0.783 | 0.7549 | 7.73 |

values of ISS are compared. In discrimination power, the models with estimated ISS perform marginally better. It appears that ISS alone is not as effective an outcome predictor as the clinical inputs are collectively. The threshold values at which equal specificity and sensitivity are achieved are also provided. In the younger age group, the thresholds are in good agreement with the value of 16 in the one definition of major trauma, but are somewhat lower in the older age group.

## 5.10 TRISS-like Models

Having derived our own models for RTS and ISS, we next proceed to incorporate them into the logistic regression framework of TRISS. It will be of interest to determine whether the TRISS formulation can improve upon the performance of the *GNOSIS* classifier developed above, in which the clinical inputs were mapped directly into survival outcome. It is interesting to note that if ISS is interpreted as a hidden node, TRISS takes a PNN tact in that it utilizes the output of that node to enhance the model output.

In computing TRISS probabilities for survival outcome, we examined three different RTS and ISS combinations, namely:

- RTS with Champion-Sacco coefficients and actual ISS values
- RTS with coefficients fitted using UVA data and actual ISS values
- RTS with coefficients fitted using UVA data and PNN-estimated ISS values.

The RTS coefficients fitted using UVA data (Table 8) were with respect to the two age groups. Logistic regressions of RTS and ISS against survival outcome were performed separately for the two age groups and also separately for blunt and penetrating injury cases, for a total of twelve models (three RTS/ISS computation methods times two age groups times two categories for B/P). Since the coded AGE values were uniform in each of the twelve sets upon which logistic regression models were derived, the AGE term was subsumed into the constant term. For the upper AGE group with blunt trauma, for instance, the constant term is taken as $\theta_0^{\text{CS}} = -1.2470 - 1.9052 \times 1 = -3.1522$. The sizes of the latter four B/P and AGE categories are provided in Table 12.

In each of the twelve cases, the coefficients obtained from logistic regression on the UVA data are compared to the Champion-Sacco (denoted by the superscript 'CS') TRISS coefficient values in

**Table 12: Age and B/P Subgroups**

| | AGE $< 55$ | AGE $\geq 55$ | Total |
|---|---|---|---|
| B | 2,329 | 892 | 3,221 |
| P | 370 | 37 | 407 |
| Total | 2,699 | 929 | 3,628 |

Table 13. $\widehat{\text{ISS}}$ denotes PNN-estimated ISS values, while unadorned ISS denotes actual values. The coefficient values determined using the UVA data are in fair agreement with the Champion-Sacco values only in the $\text{RTS}^{\text{CS}}$/ISS models for blunt trauma (i.e., the top two rows). Correspondence in the penetrating injury groups is much fainter, and the resemblance to the Champion-Sacco values disappears altogether in the $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ models, where the signs of the coefficients are no longer even the same.

**Table 13: TRISS Coefficients**

| B/P | Ages | RTS/ISS Values | $\theta_0^{\text{CS}}$ | $\theta_{\text{RTS}}^{\text{CS}}$ | $\theta_{\text{ISS}}^{\text{CS}}$ | $\theta_0^{\text{UVA}}$ | $\theta_{\text{RTS}}^{\text{UVA}}$ | $\theta_{\text{ISS}}^{\text{UVA}}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | $< 55$ | $\text{RTS}^{\text{CS}}$/ISS | $-1.2470$ | 0.9544 | $-0.0768$ | $-0.2822$ | 0.9032 | $-0.0972$ |
| | $\geq 55$ | $\text{RTS}^{\text{CS}}$/ISS | $-3.1522$ | 0.9544 | $-0.0768$ | $-3.3209$ | 0.9405 | $-0.0690$ |
| 1 | $< 55$ | $\text{RTS}^{\text{CS}}$/ISS | $-0.6024$ | 1.1430 | $-0.1516$ | $-19.4863$ | 11.7949 | $-1.4319$ |
| | $\geq 55$ | $\text{RTS}^{\text{CS}}$/ISS | $-3.2700$ | 1.1430 | $-0.1516$ | $-7.5218$ | 1.5021 | $-0.1536$ |
| 0 | $< 55$ | $\text{RTS}^{\text{UVA}}$/ISS | $-1.2470$ | 0.9544 | $-0.0768$ | 0.4805 | 0.6726 | $-0.0957$ |
| | $\geq 55$ | $\text{RTS}^{\text{UVA}}$/ISS | $-3.1522$ | 0.9544 | $-0.0768$ | $-7.1667$ | 0.7565 | $-0.0638$ |
| 1 | $< 55$ | $\text{RTS}^{\text{UVA}}$/ISS | $-0.6024$ | 1.1430 | $-0.1516$ | $-7.4335$ | 6.6280 | $-1.0702$ |
| | $\geq 55$ | $\text{RTS}^{\text{UVA}}$/ISS | $-3.2700$ | 1.1430 | $-0.1516$ | $-9.9043$ | 0.9561 | $-0.1434$ |
| 0 | $< 55$ | $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ | $-1.2470$ | 0.9544 | $-0.0768$ | 0.0973 | $-0.0866$ | 0.6731 |
| | $\geq 55$ | $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ | $-3.1522$ | 0.9544 | $-0.0768$ | $-11.0636$ | 0 | 0.9749 |
| 1 | $< 55$ | $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ | $-0.6024$ | 1.1430 | $-0.1516$ | $-3.0759$ | $-0.1354$ | 1.2728 |
| | $\geq 55$ | $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ | $-3.2700$ | 1.1430 | $-0.1516$ | $-11.5036$ | $-0.1139$ | 1.0771 |

ROC curve, log-likelihood, and goodness-of-fit statistics are provided in Table 14. In computing output probabilities, Champion-Sacco coefficients were used in the $\text{RTS}^{\text{CS}}$/ISS models, whereas coefficients fitted based on the UVA data were used in the $\text{RTS}^{\text{UVA}}$/ISS and $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ models. The numbers show that if actual values of ISS are used retrospectively, classification performances in all four B/P and AGE categories are virtually indifferent to whether Champion-Sacco or UVA-fitted coefficients are used. In both cases, discrimination power is excellent in the younger age group for both blunt and penetrating injuries. Discrimination power in the older group is poorer. The same is true in the $\text{RTS}^{\text{UVA}}/\widehat{\text{ISS}}$ models, although the discrimination power in all four subgroups drops slightly. As a pre-hospital tool, it is evident that this TRISS-like algorithm performs remarkably well in the younger group. The older group notwithstanding, the model performance is competitive with the 95.5% benchmark of the *GNOSIS* classifier.

**Table 14: TRISS Model Performance Statistics**

| B/P | Ages | RTS/ISS Values | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| 0 | $< 55$ | $RTS^{CS}/ISS$ | 0.9337 | 0.981 | 586.3 | 233.6 | 0.398 | 0.6440 |
| | $\geq 55$ | $RTS^{CS}/ISS$ | 0.7713 | 0.831 | 418.4 | 289.5 | 0.692 | 0.0558 |
| 1 | $< 55$ | $RTS^{CS}/ISS$ | 0.9944 | 1.000 | 131.8 | 22.3 | 0.169 | – |
| | $\geq 55$ | $RTS^{CS}/ISS$ | 0.8333 | 0.909 | 32.8 | 12.1 | 0.369 | 0.9144 |
| 0 | $< 55$ | $RTS^{UVA}/ISS$ | 0.9375 | 0.981 | 586.3 | 231.6 | 0.395 | 0.6367 |
| | $\geq 55$ | $RTS^{UVA}/ISS$ | 0.7686 | 0.829 | 418.4 | 286.8 | 0.685 | 0.0913 |
| 1 | $< 55$ | $RTS^{UVA}/ISS$ | 0.9972 | 1.000 | 131.8 | 5.4 | 0.040 | – |
| | $\geq 55$ | $RTS^{UVA}/ISS$ | 0.8333 | 0.909 | 32.8 | 10.0 | 0.305 | 0.7612 |
| 0 | $< 55$ | $RTS^{UVA}/\widehat{ISS}$ | 0.9224 | 0.981 | 586.3 | 255.4 | 0.436 | 0.6367 |
| | $\geq 55$ | $RTS^{UVA}/\widehat{ISS}$ | 0.6963 | 0.776 | 418.4 | 300.1 | 0.717 | – |
| 1 | $< 55$ | $RTS^{UVA}/\widehat{ISS}$ | 0.9915 | 0.997 | 131.8 | 18.6 | 0.141 | 0.9860 |
| | $\geq 55$ | $RTS^{UVA}/\widehat{ISS}$ | 0.8333 | 0.954 | 32.8 | 11.4 | 0.347 | 0.1640 |

## 5.11 Generalization of TRISS via PNN Models

The TRISS models in the previous section were all logistic regression models of RTS and ISS against survival outcome. With classification PNNs, however, we can abandon the restrictive assumptions of conventional logistic regression, i.e., the linear algebraic form of the logit polynomials, and synthesize a more flexible and accurate model of survival probability as a function of the clinical inputs (as used in the previous *GNOSIS* classifier) and ISS (actual or estimated). Table 15 displays the results of *GNOSIS* models with 2nd-degree nodal polynomials that give survival probability as functions of AGE, B/P, EY, VB, MT, RR, and SBP.

**Table 15: Generalized TRISS Model Performance Statistics**

| ISS Values | Ages | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|---|
| Actual | all | 0.8935 | 0.946 | 1198.7 | 499.8 | 0.417 | 0.0001 |
| | $< 55$ | 0.9565 | 0.975 | 720.6 | 186.3 | 0.258 | 0.0001 |
| | $\geq 55$ | 0.8226 | 0.901 | 455.4 | 248.4 | 0.545 | 0.0425 |
| Estimated | all | 0.8911 | 0.952 | 1198.7 | 527.4 | 0.440 | 0.0001 |
| | $< 55$ | 0.9500 | 0.973 | 720.58 | 219.0 | 0.304 | 0.0040 |
| | $\geq 55$ | 0.8213 | 0.887 | 455.4 | 259.80 | 0.570 | 0.2437 |

From the results in Table 15, one observes that the performance of the models with actual ISS values are marginally better than that of the models incorporating estimated ISS values. Moreover, the performance with the estimated ISS values are virtually identical to those of the earlier *GNOSIS* classifier models with 2nd-degree nodal polynomials. Henceforth, we shall refer to the model for all ages using actual ISS values as Model II. Whereas Model I is our main result for pre-hospital use, Model II represents our penultimate result for *ex post* evaluation of hospitalized patients.

In the young age group, there is a slight gain in discrimination power (95.0% vs. 94.4%), but the impression overall suggests that use of estimated ISS as, in effect, a hidden layer does not contribute significantly to classification performance. However, the slight edge in both age groups and in the full database implies that use of ISS in pre-hospital algorithms would offer an incremental gain were it possible to infer ISS directly.

## 5.12 Utilization of AIS Scores

As the final set of analyses for survival prediction with the UVA data, we broke down the final remnant of artifice in the conventional scores, namely the manner in which ISS is a composition of AIS scores that measure the severity of injury in specific body regions. That ISS is the sum of the squares of the three highest AIS scores seems, on face value, arbitrary and *ad hoc*. PNN synthesis methods naturally beckon the freedom to do as they see fit with the six AIS scores, rather than have them summarized and prefiltered *a priori* via the ISS convention.

An alternative scoring system takes an almost identical approach to TRISS in modeling survival probability, except that it uses AIS scores rather than ISS. Like TRISS, ASCOT models survival outcome as a logistic regression model in standard physiological and anatomical criteria, as described in Appendix A. Instead of ISS alone, it examines three regions, the variables for which are denoted as A, B, and C. Based on the description in the literature and presented in Appendix A, we interpreted A to correspond to AIS region 1, B to AIS region 3, and C to the other four AIS regions. With these assumptions, we performed logistic regressions on the UVA data and compared the resulting coefficient values with those cited in the literature. Results are given in Table 16. Clearly, numerical agreement between the official, MTOS-based ASCOT coefficient values and the values that we obtained by fitting on UVA data is not good at all; about all that can be said is that the signs agree. Classification performance statistics are given in Table 17. The discrimination power is very good for penetrating injuries, but mediocre for blunt cases. This suggests that superior classification results in penetrating injury cases can be obtained through incorporation of more refined anatomical data. We attempted to obtain better definition of anatomical injury by appealing to the six AIS codes and comparing results with singular use of ISS. We first constructed a survival predictor by training *GNOSIS* on the six AIS scores, the results of which are given in

**Table 16: ASCOT Coefficients**

| | Blunt | | Penetrating | |
|---|---|---|---|---|
| | MTOS Values | UVA Values | MTOS Values | UVA Values |
| $\theta_0$ | −1.1570 | −2.9785 | −1.1350 | −10.5482 |
| $\theta_{\text{GCS}}$ | 0.7705 | 0.9322 | 1.0626 | 3.6849 |
| $\theta_{\text{SBP}}$ | 0.6583 | 0.9633 | 0.3638 | 0.6255 |
| $\theta_{\text{RR}}$ | 0.2810 | 0.3311 | 0.3332 | 1.9187 |
| $\theta_{\text{A}}$ | −0.3002 | −1.2886 | −0.3702 | −2.4271 |
| $\theta_{\text{B}}$ | −0.1961 | −0.4047 | −0.2053 | 5.5846 |
| $\theta_{\text{C}}$ | −0.2086 | −0.4679 | −0.3188 | −8.8582 |
| $\theta_{\text{AGE}}$ | −0.6355 | −2.0591 | −0.8365 | −10.6732 |

**Table 17: ASCOT Model Performance Statistics**

| B/P | Coefficients | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|---|
| 0 | MTOS | 0.8524 | 0.9402 | 1025.1 | 713.5 | 0.696 | 0.0006 |
| | UVA | 0.8482 | 0.9400 | 1025.1 | 552.5 | 0.540 | 0.0343 |
| 1 | MTOS | 0.9570 | 0.9939 | 171.2 | 68.3 | 0.399 | 0.0006 |
| | UVA | 0.9632 | 0.9966 | 171.2 | 21.4 | 0.125 | 0.0855 |

Table 18, where a classification PNN with 2nd-degree nodal polynomials was used. The discrimi-

**Table 18: AIS-based Survival Predictor**

| Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|
| 0.8361 | 0.900 | 1198.7 | 623.47 | 0.520 | 0.0005 |

nation power of 83.6% represents a slight improvement over the 82.1% obtained with ISS alone as a threshold classifier.

We next synthesized estimation models for each of the AIS scores, with standard clinical inputs AGE, B/P, EY, VB, MT, RR, and SBP, as before. Third-degree nodal polynomials were used. RMS estimation errors are compared to the standard deviation of the AIS distributions in the database. Results are given in Table 19. These results are poor in that the RMS estimation

**Table 19: AIS Estimation Models**

| | RMS Estimation Error | Standard Deviation |
|---|---|---|
| $AIS_1$ | 1.0587 | 1.4022 |
| $AIS_2$ | 0.5107 | 0.5293 |
| $AIS_3$ | 0.9630 | 1.0670 |
| $AIS_4$ | 0.8228 | 0.8687 |
| $AIS_5$ | 1.1770 | 1.2372 |
| $AIS_6$ | 0.7749 | 0.8346 |

errors are only marginally less than the intrinsic standard deviations of the AIS distributions in the database. The individual AIS values cannot readily be inferred from examination of the clinical indicators. This makes sense heuristically since the standard clinical indicators pertain to overall medical condition and seldom necessarily reflect specific regions of the body. To examine the use of the AIS scores, rather then ISS, we trained classifier models on the standard clinical inputs and the six AIS scores, once with actual values and again with estimated values. Results are given in Table 20. The numbers demonstrate an improvement in discrimination power for the full database over the *GNOSIS* classifier that used only the clinical inputs (92.3% vs. 90.8%). Since the *estimated* AIS

**Table 20: Survival Predictor based on Clinical Inputs and AIS Scores**

| Model | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|
| Actual AIS Scores | 0.9225 | 0.975 | 1198.7 | 356.09 | 0.297 | 0.0001 |
| Estimated AIS Scores | 0.9036 | 0.906 | 1198.7 | 465.23 | 0.388 | 0.0001 |

values do not represent any such improvement, however, it follows that this tool cannot presently serve as a pre-hospital triage algorithm. It does demonstrate, however, that greater knowledge of anatomical injury would definitely improve the quality of pre-hospital decisions.

The results in Table 20 constitute our ultimate model for *ex post* use (Model III), which was synthesized with training on the seven clinical inputs and the six actual AIS scores, for a total of thirteen inputs. This differs from Model II in that the AIS components are treated individually, rather than funneled through ISS. Performance of Model III over the three age groups is given in Table 21. The results in Table 21 demonstrate considerably better performance than both Models

**Table 21: Generalized ASCOT Performance**

| Ages | Discrimination Power | Area Under ROC Curve | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ | $p$ |
|---|---|---|---|---|---|---|
| all | 0.9225 | 0.9698 | 1198.7 | 356.0 | 0.297 | 0.0001 |
| $< 55$ | 0.9847 | 0.9979 | 720.6 | 96.3 | 0.134 | – |
| $\geq 55$ | 0.8746 | 0.9868 | 455.4 | 290.8 | 0.639 | – |

I and II. It therefore represents substantial improvement over the state-of-the-art standards (i.e., TRISS and ASCOT) for *ex post* evaluation of survival outcomes among groups of patients with common injury attributes and assessing the quality of care received. We have thus demonstrated the superiority of the PNN models over existing scoring systems for both pre-hospital and *ex post* evaluation of hospitalized patients.

## 5.13 Trichotomous Classification Analysis of UVA Data

As the final set of analyses on the UVA data, we sought to perform a trichotomous classification of the UVA database exemplars into RED, AMBER, and GREEN categories, as explicitly envisioned in the original solicitation (see Section 1. For this objective, we presented *GNOSIS* with a training database comprised of all seven standard clinical inputs and three class outputs, which we defined as follows:

- RED constitutes all nonsurvivors
- AMBER constitutes all survivors with actual ISS $\geq 16$
- GREEN constitutes all survivors with actual ISS $< 16$

The survivor category is thus partitioned into those having and those not having sustained major trauma in the one conventionally-defined sense. Such a refinement facilitates focus on those living patients in need of critical care services and should help identify distinguishing attributes common to such cases.

We trained a *GNOSIS* classification PNN with 2nd-degree nodal polynomials on the full UVA database. The outputs of the resulting PNN are a triplet of class membership probabilities ($\pi_{\text{RED}}$, $\pi_{\text{AMBER}}$, $\pi_{\text{GREEN}}$) that sum to unity. To use these probabilities to construct decisional algorithms, however, is much more complicated in the trichotomous than in the dichotomous case. Whereas the decision rule in the latter is simply a matter of whether the value of the one output probability degree of freedom exceeds a threshold, the decisional output in the trichotomous case is a function of two probability degrees of freedom. Fig. 6 illustrates the nature of such a decision rule in graphical form. The space of the two probability degrees of freedom (say $\pi_{\text{RED}}$ and $\pi_{\text{GREEN}}$) is a triangular

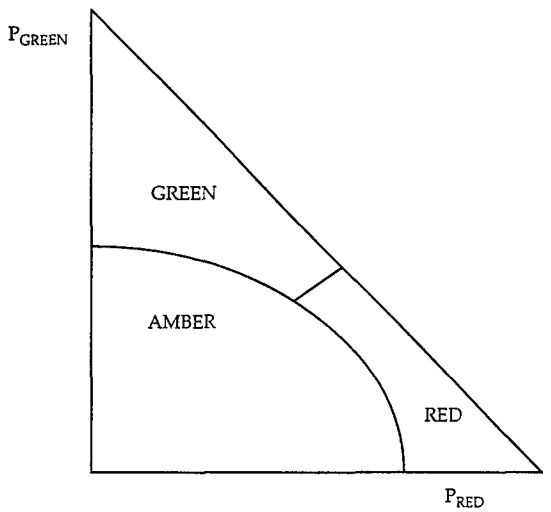


**Figure 6: Decisional Surface for Trichotomous Classification**

region, since their sum is constrained to be less than unity. With $\pi_{\text{RED}}$ and $\pi_{\text{GREEN}}$ on the horizontal and vertical axes respectively, decisional outputs are obtained by partitioning the triangular region in a suitable fashion, such as shown.

Although there are infinitely many ways of constructing such a decisional partition, it is still possible to construct a three-dimensional ROC surface, in which the diagonal elements of the classification performance matrix, $\Pi$, are the plot axes. For a given decisional partition scheme for the triangular region in Fig. 6, certain values for $\Pi_{\text{R,R}}$, $\Pi_{\text{A,A}}$, and $\Pi_{\text{G,G}}$ will be obtained. At that point in $\Pi_{\text{R,R}}$-$\Pi_{\text{A,A}}$-$\Pi_{\text{G,G}}$ space, the following question may be posed unambiguously: Is it possible to find a different partition of the triangular region such that one or more of the diagonal $\Pi$ elements can be increased without reducing the others? The manifold of realizable points at which the answer is "no" constitutes a three-dimensional ROC surface.

Since the construction of such ROC surfaces would be prohibitively time-consuming, we took a more modest approach to assessing the performance characteristics of potential decision rules for the trichotomous problem. As the first step, we created a threshold to segregate AMBERs from non-AMBERs; this resulted in a pair of specificity/sensitivity characteristics as a function of that threshold. Among the group of putative non-AMBERs at each such threshold level, we then measured the discrimination power (found using a second threshold) between the REDs and GREENs. Results are given in Table 22. The specificities and sensitivities for the AMBER/non-AMBER decision are shown in the middle two columns. The discrimination power, $\Delta$, between REDs and GREENs achievable at the given $\xi_{\text{AMBER}}$ setting is tabulated in the fourth column.

**Table 22: Classification Performance for Trichotomous Problem**

| $\xi_{\text{AMBER}}$ | $\Pi_{\text{not A, not A}}$ | $\Pi_{\text{A,A}}$ | $\Delta$ |
|---|---|---|---|
| 0.05 | 0.1737 | 0.9504 | 0.9781 |
| 0.10 | 0.6702 | 0.7087 | 0.9268 |
| 0.15 | 0.8467 | 0.5702 | 0.9054 |
| 0.20 | 0.9094 | 0.4711 | 0.8954 |
| 0.25 | 0.9335 | 0.4194 | 0.8877 |
| 0.30 | 0.9504 | 0.3740 | 0.8810 |
| 0.35 | 0.9602 | 0.3285 | 0.9162 |
| 0.40 | 0.9660 | 0.2934 | 0.9133 |
| 0.45 | 0.9749 | 0.2665 | 0.9089 |
| 0.50 | 0.9809 | 0.2355 | 0.9063 |
| 0.55 | 0.9870 | 0.1963 | 0.9009 |
| 0.60 | 0.9911 | 0.1529 | 0.8961 |
| 0.65 | 0.9940 | 0.1281 | 0.8924 |
| 0.70 | 0.9965 | 0.1074 | 0.9130 |
| 0.75 | 0.9975 | 0.0909 | 0.9118 |
| 0.80 | 0.9994 | 0.0537 | 0.9086 |
| 0.85 | 1.0000 | 0.0310 | 0.9069 |
| 0.90 | 1.0000 | 0.0083 | 0.9064 |
| 0.95 | 1.0000 | 0.0000 | 0.9062 |
| 1.00 | 1.0000 | 0.0000 | 0.9062 |

Between the AMBERs and non-AMBERs, discrimination power of only 69% is obtained at a $\xi_{\text{AMBER}}$ setting just above 0.10. At that setting, the discrimination power between REDs and GREENs in the non-AMBER group is 92.7%. This suggests that detecting AMBERs, who lie in the gray region between life and death, is considerably more difficult than distinguishing survivors from nonsurvivors.

Results for the other two alternatives are given in Table 23. The results for these alternative decision methods are no better. For RED as the first decision, discrimination power of about 89% is achieved at a threshold below 0.05. The discrimination power between AMBER and GREEN at that level is about 70%; not good. For GREEN as the first decision, discrimination power of about 71% is achieved at a threshold between 0.85 and 0.90. The discrimination power between AMBER and RED at that level is about 89%. There is thus a tradeoff in the discrimination powers after the two decisions, but AMBER as the first decision seems to be the most effective type of decisional partition. Additional medical data on patients would be required to improve identification of the AMBER group.

# 6 Algorithm Refinement with North Carolina Data

For the Phase I project, we acquired two additional databases for further studies, namely the North Carolina state Trauma Registry and an ICU database. These data sets, both provided to BAI

**Table 23: Classification Performance for Trichotomous Problem**

| $\xi_{RED}$ | $\Pi_{not\ R,\ not\ R}$ | $\Pi_{R,R}$ | $\Delta$ | $\xi_{GREEN}$ | $\Pi_{not\ G,\ not\ G}$ | $\Pi_{G,G}$ | $\Delta$ |
|---|---|---|---|---|---|---|---|
| 0.05 | 0.9188 | 0.8592 | 0.7005 | 0.05 | 0.1470 | 0.9997 | 0.8286 |
| 0.10 | 0.9555 | 0.7958 | 0.6610 | 0.10 | 0.2284 | 0.9977 | 0.7910 |
| 0.15 | 0.9676 | 0.7606 | 0.6782 | 0.15 | 0.3115 | 0.9950 | 0.7679 |
| 0.20 | 0.9722 | 0.7324 | 0.6882 | 0.20 | 0.3435 | 0.9937 | 0.7500 |
| 0.25 | 0.9785 | 0.6972 | 0.6989 | 0.25 | 0.3546 | 0.9930 | 0.7660 |
| 0.30 | 0.9831 | 0.6197 | 0.7143 | 0.30 | 0.3674 | 0.9923 | 0.7651 |
| 0.35 | 0.9862 | 0.5775 | 0.7228 | 0.35 | 0.3738 | 0.9917 | 0.7532 |
| 0.40 | 0.9905 | 0.5493 | 0.7000 | 0.40 | 0.3754 | 0.9910 | 0.7580 |
| 0.45 | 0.9925 | 0.5070 | 0.7068 | 0.45 | 0.3898 | 0.9893 | 0.7706 |
| 0.50 | 0.9943 | 0.4648 | 0.7122 | 0.50 | 0.4042 | 0.9843 | 0.7668 |
| 0.55 | 0.9957 | 0.3944 | 0.7199 | 0.55 | 0.4217 | 0.9797 | 0.7650 |
| 0.60 | 0.9968 | 0.3803 | 0.7219 | 0.60 | 0.4489 | 0.9747 | 0.7837 |
| 0.65 | 0.9980 | 0.3451 | 0.7263 | 0.65 | 0.4760 | 0.9680 | 0.7885 |
| 0.70 | 0.9989 | 0.3169 | 0.7296 | 0.70 | 0.5144 | 0.9594 | 0.8135 |
| 0.75 | 0.9989 | 0.2958 | 0.7310 | 0.75 | 0.5447 | 0.9367 | 0.8220 |
| 0.80 | 0.9991 | 0.2676 | 0.7333 | 0.80 | 0.5863 | 0.9044 | 0.8396 |
| 0.85 | 0.9994 | 0.2465 | 0.7351 | 0.85 | 0.6645 | 0.8334 | 0.8608 |
| 0.90 | 0.9997 | 0.2394 | 0.7360 | 0.90 | 0.7764 | 0.6402 | 0.8863 |
| 0.95 | 1.0000 | 0.1197 | 0.7438 | 0.95 | 0.9617 | 0.1436 | 0.8922 |

by Dr. Robert Rutledge of the University of North Carolina at Chapel Hill, shall herein be referred to as NCTR and NCICU respectively. The former contained pre-hospital and ER data in all of the standard UVA data fields (AGE, B/P, EY, VB, MT, RR, SBP, ISS, and survival outcome) plus three additional fields: heart rate (HR), body temperature (T), and the hematocrit ratio (HCT). The NCICU database provided acute physiological data and most of the standard clinical fields for hospitalized patients on a day-by-day basis; this provided an opportunity to explore time-series modeling methods to enhance the inference and prediction methods explored and developed thus far. The NCTR data served three different purposes: (1) transdatabase comparison of models derived using the UVA database; and (2) assessing model performance improvements accruing to the three additional data fields; and (3) assessing approaches to deal with missing input data fields.

## 6.1 Overview of NCTR Data

The NCTR database provided all of the standard clinical inputs that were available in the UVA data, along with ISS and survival outcome. It also provided heart rate (in beats per minute), body temperature (in °F), and the hematocrit. The NCTR database contained 4,125 complete patient records, with babies (< 2 years old) and extremely old people (> 95 years old) excluded. Among the NCTR patients, 472 (11.4%) had penetrating injuries and 146 (3.5%) died, distributions very similar to those in the UVA data. Univariate histogram distributions with respect to each of the input variables, including the three new ones, are shown in Fig. 7. Summary statistics are provided in Table 24.

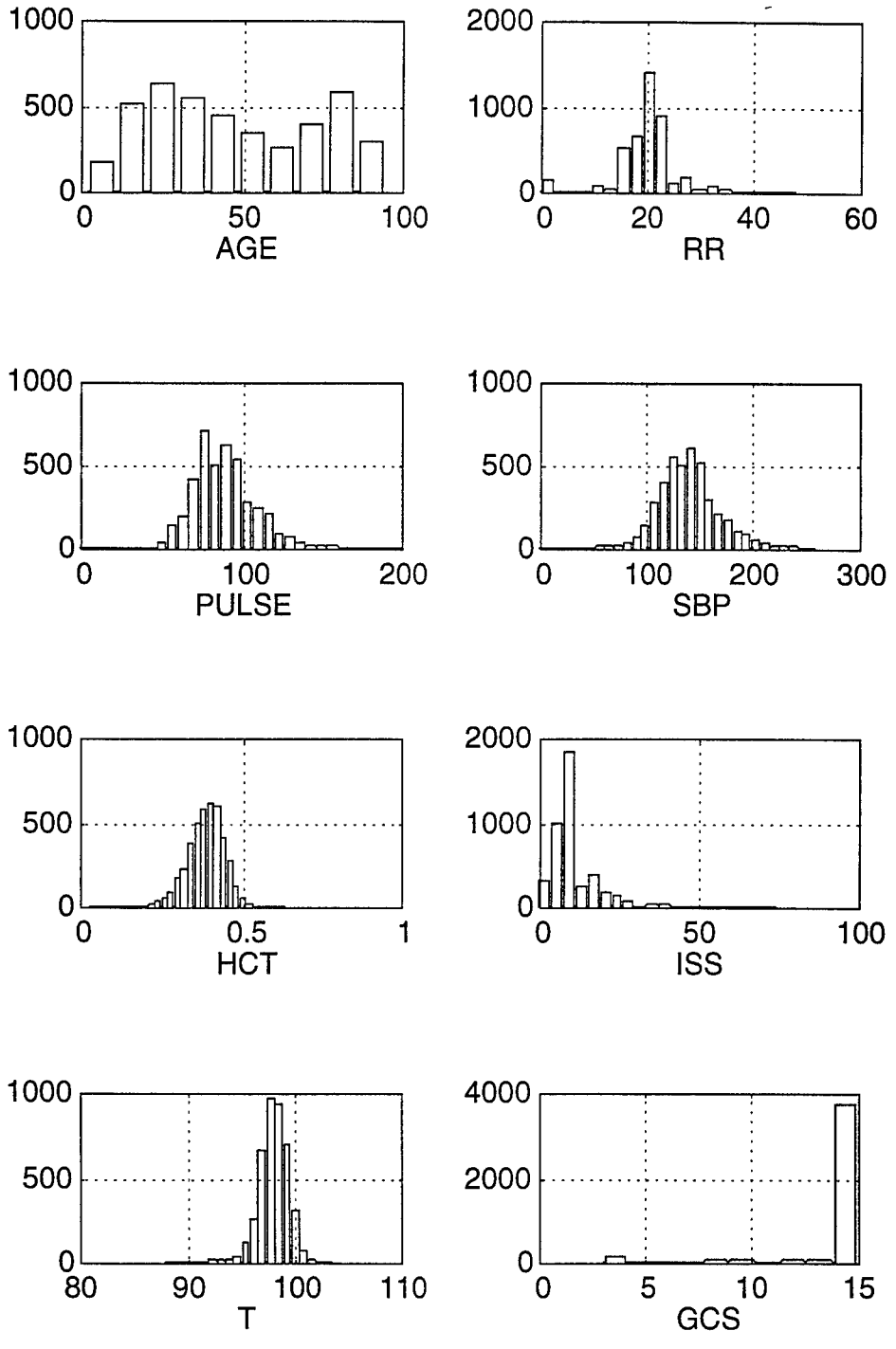


**Figure 7: Univariate Distributions for the NCTR Database**

**Table 24: NCTR Univariate Distribution Statistics**

| Field | Mean ± Std. Dev. |
|---|---|
| T | 98.04 ± 1.40 |
| RR | 20.26 ± 6.61 |
| HR | 89.46 ± 19.83 |
| SBP | 140.74 ± 28.43 |
| HCT | 0.39 ± 0.06 |
| ISS | 10.10 ± 7.21 |

## 6.2 Transdatabase Comparison Tests

The availability of two completely independent trauma databases, constructed and administered by two separate institutions in different states, provided a valuable opportunity to test the classification performance of the UVA-derived models on the NCTR database, and vice versa. Such comparisons are important because they furnish stringent tests of the true *generality* of the models.

As the first set of comparison tests, we trained separate *GNOSIS* classification models for survival outcome on the two databases, UVA and NCTR. The resulting UVA-derived and NCTR-derived models were then self- and cross-validated (i.e., evaluating the UVA-derived model on the NCTR data, and vice versa). ROC curves were obtained, the salient performance characteristics of which are given in Table 25. The results indicate poorer evaluation performance on NCTR data

**Table 25: Self-Validated and Cross-Validated Classification Performance**

| Training Database | Evaluation Database | Discrimination Power | Area Under ROC Curve | $\xi$ | $-2\Lambda_0$ | $-2\Lambda$ | $\Lambda/\Lambda_0$ |
|---|---|---|---|---|---|---|---|
| UVA | UVA | 0.8907 | 0.9517 | 0.0335 | 1198.7 | 555.4 | 0.463 |
| UVA | NCTR | 0.7628 | 0.8175 | 0.0256 | 1262.4 | 1067.2 | 0.845 |
| NCTR | UVA | 0.8629 | 0.9375 | 0.0326 | 1198.7 | 676.3 | 0.564 |
| NCTR | NCTR | 0.7863 | 0.8743 | 0.0327 | 1262.4 | 850.2 | 0.436 |

than on the UVA data. The NCTR database, for unknown reasons, is inherently more difficult to model than the UVA database. The classification performance on a particular database, however, is relatively insensitive to which model is used. For instance, the discrimination power of the NCTR-derived model evaluated on the UVA data trails that of the UVA-derived model evaluated introspectively (i.e., with self-validation) by only three percentage points. Moreover, the threshold placements at which specificity and sensitivity coincide are almost identical ($\xi \approx 0.033$) in three out of the four cases.

These findings indicate reasonably good transportability of data in that reasonably consistent specificity-sensitivity characteristics, with fixed thresholds, can be obtained even with models trained on entirely different databases. This bodes well for the prospects of developing practical algorithms for use in the pre-hospital environment. Decisional performance is much more dependent on the nature of the evaluation database, or the idiosyncracies of the trauma care environment, than on the particular patient population from which the models were derived.

As a second set of transdatabase comparison tests, we analyzed the reliability of the classification

algorithms on the two databases; results are given in Table 26. The first row in the upper table

**Table 26: Reliability Results for UVA and NCTR Databases**

| Evaluation on UVA Data | | | |
|---|---|---|---|
| Actual Outcome | Prediction of UVA-Derived Model | Prediction of NCTR-Derived Model | Incidents |
| LIVE | LIVE | LIVE | 2,954 |
| LIVE | LIVE | DIE | 144 |
| LIVE | DIE | LIVE | 84 |
| LIVE | DIE | DIE | 304 |
| DIE | LIVE | LIVE | 12 |
| DIE | LIVE | DIE | 3 |
| DIE | DIE | LIVE | 7 |
| DIE | DIE | DIE | 120 |

| Evaluation on NCTR Data | | | |
|---|---|---|---|
| Actual Outcome | Prediction of UVA-Derived Model | Prediction of NCTR-Derived Model | Incidents |
| LIVE | LIVE | LIVE | 3,010 |
| LIVE | LIVE | DIE | 228 |
| LIVE | DIE | LIVE | 133 |
| LIVE | DIE | DIE | 608 |
| DIE | LIVE | LIVE | 27 |
| DIE | LIVE | DIE | 18 |
| DIE | DIE | LIVE | 5 |
| DIE | DIE | DIE | 96 |

indicates that in the UVA database, there are 2,954 surviving patients who were identified correctly by both the UVA-derived and NCTR-derived models as surviving. The decisional outputs in Table 26 were based on a threshold setting of $\xi = 0.033$ in both models.

For evaluation on the UVA data, the specificity-sensitivity was 88.9%-89.4% for the model derived using the UVA data and 87.1%-86.6% for the model derived using the NCTR data. The corresponding results for evaluation on the NCTR data were 81.4%-69.2% and 79.0%-78.1%, respectively. The reliability statistics, however, are the rates at which decisional outputs are correct. For a given decisional model and evaluation database, the reliability indicators $R_{\mathrm{n,n}}$ and $R_{\mathrm{p,p}}$ are respectively the percentages of negative and positive decisions that are correct. For evaluation on the UVA data, the model derived using the UVA data has reliability indicator scores of

$$R_{\mathrm{n,n}} = \frac{2,954 + 144}{2,954 + 144 + 12 + 3} = 0.995 \quad \text{and} \quad R_{\mathrm{p,p}} = \frac{7 + 120}{84 + 304 + 7 + 120} = 0.247 \qquad (4)$$

which indicate that a negative decision by the model derived using the UVA data is correct in 99.5% of cases, whereas a positive decision is correct in only 24.7% of cases. The corresponding statistics for the model derived based on the NCTR data but evaluated on the UVA data are $R_{\mathrm{n,n}} = 99.4\%$

and $R_{p,p} = 21.5\%$. The decisional outputs of the two models agree in 93.4% of all cases. The very low $R_{p,p}$ figures, although obviously disconcerting from the EMT point of view, are, in fact, necessary to achieve desired target values of sensitivity and specificity. To obtain good $R_{n,n}$ and $R_{p,p}$ results simultaneously would require radically different threshold settings that would result in extremely poor $\Pi_{n,n}$ or $\Pi_{p,p}$. To improve both specificity-sensitivity and reliability characteristics simultaneously would require more refined medical knowledge capable of distinguishing survivor and nonsurvivor attributes more accurately.

Results for evaluation on the NCTR database are even worse, with $R_{p,p}$ values of 12.0% for models derived using both the UVA and NCTR databases. One way to improve these results is to merge the decisional outputs of the two models, such that a positive finding is declared if and only if both models declare positive findings. The statistics then improve slightly to $R_{p,p} = 28.3\%$ for evaluation on the UVA data and $R_{p,p} = 13.6\%$ for evaluation on the NCTR data, without appreciable diminution of specificity-sensitivity characteristics.

## 6.3 Inclusion of T, HR, and HCT Inputs

The second major purpose of analyzing the NCTR data was to assess improvements in model performance accruing to the additional input fields (T, HR, and HCT) that were available in the NCTR database but not the UVA database. Univariate correlation with ISS and survival outcome are tabulated in Table 27. All three have "normal" ranges in which the ISS values are generally

**Table 27: Univariate Analysis of HR, T, and HCT Distributions in NCTR Database**

| T | mean ISS |
|---|---|
| < 92 | 20.73 |
| 92-94 | 19.32 |
| 94-96 | 13.26 |
| 96-98 | 10.07 |
| 98-100 | 9.37 |
| 102 | 12.44 |

| P | mean ISS |
|---|---|
| < 50 | 15.16 |
| 50-75 | 9.79 |
| 75-100 | 9.29 |
| 100-125 | 10.88 |
| 125-150 | 15.16 |
| > 150 | 17.77 |

| HCT | mean ISS |
|---|---|
| < 0.3 | 13.37 |
| 0.3-0.35 | 11.08 |
| 0.35-0.4 | 10.01 |
| 0.4-0.45 | 9.50 |
| 0.45-0.5 | 9.10 |
| > 0.5 | 11.09 |

below average. Above- and below-normal deviations are both associated with above-average ISS values. This pattern, in the RR and SBP fields, was similarly observed in the UVA data.

*GNOSIS* was trained on the NCTR database to examine the performance improvements accruing to the new inputs. ISS estimation models with second-degree polynomial nodal elements in all ten inputs (AGE, B/P, SBP, RR, T, HR, HCT, EY, VB, and MT) were synthesized and compared with and without inclusion of T and HCT. The RMS estimation errors, upon completion of each layer (with projection pursuit), are provided in Table 28. A slight, but noticeable, improvement is achieved by inclusion of T and HCT in the estimation model for ISS. The loss of such an input value (e.g., due to a certain biomedical instrument not being available or operative) does result in significant diminution of estimation or classification performance. Use of higher-order nodal polynomials, as with the UVA data, improves performance substantially, as illustrated in Table 29. It is noteworthy that the best-case RMS estimation error matches closely that for the UVA data, even though the two databases are completely independent, but have similar distributions of patients. In both cases, *GNOSIS* produced models that outperformed conventional regression by a

**Table 28: *GNOSIS* Performance Improvements with T and HCT**

| Synthesized Layers | T and HCT excluded | T excluded HCT included | T and HCT included |
|---|---|---|---|
| 1 | 6.299 | 6.281 | 6.278 |
| 2 | 6.288 | 6.247 | 6.225 |
| 3 | 6.280 | 6.236 | 6.212 |
| 4 | 6.275 | 6.225 | 6.189 |

**Table 29: *GNOSIS* Performance Improvements with Higher-Order Polynomials**

| Synthesized Layers | Nodal Polynomial Degree | | |
|---|---|---|---|
| | 2nd-order | 3rd-order | 4th-order |
| 1 | 6.278 | 6.178 | 6.112 |
| 2 | 6.225 | 6.095 | 5.996 |
| 3 | 6.212 | 6.052 | 5.940 |
| 4 | 6.189 | 6.023 | 5.914 |

wide margin.

## 6.4 Inclusion and Exclusion of Clinical Inputs

The final analysis performed on the NCTR data was to assess approaches for dealing with missing input data fields. Using the NCTR data, we serially omitted one of the ten inputs and trained a neural network classifier using the other nine. This was done for each input variable, and the resulting ten models were evaluated. The results this way were compared to using the ten-input classifier model and inserting the average value for that variable, as computed by averaging over the entire NCTR database. In the case of the age field, certain "reasonable guesses" were made about the ease of guessing a person's age, e.g., assuming that the age of a person older than 70 can generally be estimated to within twenty years. Table 30 displays the results of the two approaches, with the discrimination power and $\Lambda$ statistics provided for comparison. The similarity of the results indicates that average values can generally be used when an input value is not readily available. Another approach to the missing data field problem demonstrated recently [76] exploits the correlation in the input data to "synthesize" missing inputs based on the data input fields that are available. Belief networks (Appendix E) furnish a potentially more elegant approach.

## 6.5 Time-Series Analysis and Dynamic Models

In many estimation problems, the underlying physical relationship between the input and output variables is fundamentally dynamical in character. In continuous time, the physical model of such systems is described by a set of differential equations, rather than by a static functional relationship. In discretized time, there are at least three distinct types of dynamical estimation models that are of great practical importance in numerous signal processing applications:

$$\underline{\hat{y}}_k = f(\underline{x}_k, \underline{x}_{k-1}, \underline{x}_{k-2}, \ldots) \tag{5a}$$

**Table 30: Model Performance with Omitted Variables**

| Omitted Variable | Neural Network Trained without Omitted Variable | | Average Value of Omitted Variable Used in Ten-Input Model | |
|---|---|---|---|---|
| | Discrimination Power | $\Lambda$ | Discrimination Power | $\Lambda$ |
| AGE | 0.7465 | 868.6 | 0.7786 | 792.5 |
| B/P | 0.8005 | 795.6 | 0.7968 | 807.3 |
| T | 0.7943 | 795.5 | 0.7863 | 805.0 |
| RR | 0.7979 | 793.2 | 0.7858 | 827.4 |
| P | 0.7856 | 810.8 | 0.7861 | 825.9 |
| SBP | 0.7911 | 796.9 | 0.7988 | 815.3 |
| EY | 0.7976 | 789.0 | 0.7554 | 946.5 |
| VB | 0.8089 | 785.6 | 0.7808 | 929.1 |
| MT | 0.8050 | 795.5 | 0.6975 | 1135.2 |
| HCT | 0.7946 | 802.0 | 0.7945 | 813.2 |

$$\underline{\hat{y}}_k = f(\underline{x}_k, \underline{x}_{k-1}, \underline{x}_{k-2}, \cdots; \underline{y}_{k-1}, \underline{y}_{k-2}, \cdots) \tag{5b}$$

$$\underline{\hat{y}}_k = f(\underline{x}_k, \underline{x}_{k-1}, \underline{x}_{k-2}, \cdots; \underline{\hat{y}}_{k-1}, \underline{\hat{y}}_{k-2}, \cdots) \tag{5c}$$

In Eqs. 5, $\underline{x}$ is a $P \times 1$ input vector and $\underline{y}$ is a $Q \times 1$ output vector. In the first equation, the estimated output value at time $k$ is a function of the present and past values of $\underline{x}$. If $f$ is a linear function, Eq. 5a is known as a *finite impulse response* (FIR) model. This is the approach that is taken below in the analysis of the NCICU data for constructing a dynamical model for survival outcome.

In Eqs. 5b and 5c, the output estimate, $\underline{\hat{y}}_k$, depends on past output values as well as the input variables. In Eq. 5b, actual past outputs are used; this is known as an *equation-error* model. If $f$ is a linear function, it reduces to an IIR (infinite impulse response) or ARMAX (autoregressive moving-average with exogenous input) model. In Eq. 5c, however, *estimated* past output values are used; this is an *output-error* model.

*Recurrent neural networks* (RNNs) are a means of fitting such models to time-series training data. *GNOSIS* can synthesize recurrent neural networks as well as the purely static PNNs that we have described throughout the present report. The resulting network models produce time-varying output signals. Recent [49] and prior [80] work by BAI has shown that RNNs can emulate real linear and nonlinear dynamical systems with high accuracy and computational efficiency. The necessary size of the training database depends on the structural validity of the surmised recurrent model. If the "correct" model structure is surmised, the model parameter values can be identified using a single time-series simulation of the system. For this reason, *a priori* knowledge about the dynamics of the underlying system, i.e., the set of differential equations governing it, can be tremendously advantageous.

Nodes in RNNs have internal shift registers that effect time delays and signal feedback. *GNOSIS*

globally optimizes the parameters in RNNs containing multiple layers of nodes with these types of feedback connections. Static PNNs, by contrast, have no such internal feedbacks or time delays in the nodes and are often called *feedforward* neural networks for this reason. In many applications, recurrent neural networks are actually much simpler than feedforward networks in architectural complexity and number of degrees of freedom. This reduces probability of overfit and increases network accuracy and robustness.

To allow the injury severity models to be updated based on the arrival of serial data updates, we sought ways in which time-series, or dynamical, models could be demonstrated. The outputs of such a model reflect information about past as well as present values of the input variables. One possibility for demonstrating such models was suggested by the North Carolina ICU (NCICU) database, in which acute physiological data, including most of the inputs we have already mentioned, were provided on a daily basis. This database, also obtained from the University of North Carolina at Chapel Hill, provided various physiological data from hospitalized patients in an intensive care unit setting. GCS, RR, SBP, and survival outcome were among the various fields provided, but ISS was not provided. The univariate distributions within the NCICU database are similar to those in UVA and NCTR, with histogram plots for RR and SBP (measured on the first day of ICU care) illustrated in Fig. 8. With respect to the principal fields examined in the UVA database, the NCICU database provided complete records for 2,152 patients.

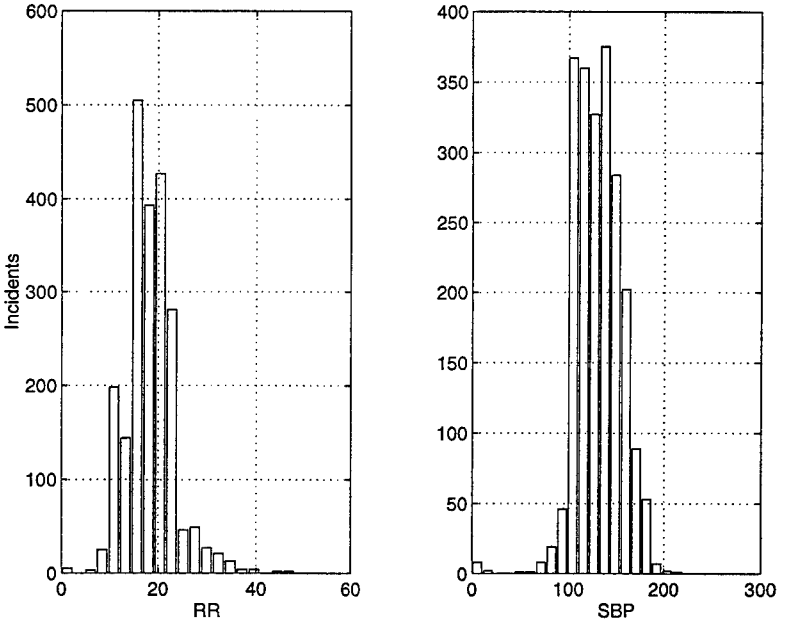


**Figure 8: Univariate Distributions for RR and SBP in NCICU Database**

GCS, as in the UVA database, was very skewed, with 1,723 out of 2,152 patients, or 80.1%, having GCS = 15. The remainder were fairly evenly divided among lower scores, with a slight preponderance at GCS=3. Among all patients, only 74 (3.4%) died.

For this effort, we sought to predict survival outcome through adaptive least squares and *finite-*

*impulse response* (FIR) models in estimated ISS. In an FIR model, one is provided with a sequence of inputs and an outputs; the output value at any given time is simply a linear combination of input values from the present and recent past. Using the NCICU data, we constructed a simple FIR model, in which the input variable was the estimated ISS on a given day (obtained using the ISS estimator synthesized using the UVA database without B/P) and the output was a revised, or *updated*, survival outcome projection. The outcome was predicted on the second day of ICU stay (1,176 of the 2,152 patients, or 54.6%, were in the ICU for at least two days) as a linear function of estimated ISS on day one and estimated ISS on day two. The FIR model

$$P_d = \frac{1}{1+e^{-L}} \tag{6a}$$

$$L = -6.2615 + 0.0442 \times \widehat{\mathrm{ISS}}_1 + 0.1652 \times \widehat{\mathrm{ISS}}_2 \tag{6b}$$

was obtained using static linear regression, in which the delayed values of ISS were provided to the regressor as if they were independent static variables. Discrimination power of 83% can be obtained this way, which is significantly better than that which was achieved (79%) using static models on the NCICU data. Further investigation of this promising area is needed in Phase II.

# 7 Conclusions of Phase I Work

We summarize Phase I accomplishments as follows:

- We have developed two neural network mortality prediction models (Models I and III) that can be used for pre-hospital triage and *ex post* evaluation of hospitalized patients respectively. Model I outperforms all of the conventional, rule-of-thumb triage scoring systems; Model III outperforms both TRISS and ASCOT.
- We have demonstrated the ability of neural networks to differentiate among traditional color-coded triage categories such as RED, AMBER, and GREEN and extended the analytic methods for classification and decisional algorithms to problems involving three or more classes.
- We have demonstrated that average values for missing input values can be used in the neural network models while achieving results not significantly different from those obtained with use of models trained without the missing variable.
- We have introduced and demonstrated one approach that was effective for time-series analysis and updating of model outputs.

Based on the success of the results outlined above, it is clear that much more can be achieved with a thorough investigation of processing algorithms (both static and dynamic model types) and more extensive (quantity), varied (military vs. civilian), and thorough (time-series) data sets. All of these will issues will be explored in detail in the Phase II effort, the proposal for which has been submitted.

# References

[1] Baker, S.P., and B. O'Neill, "The injury severity score: an update," *J. Trauma*, Vol. 16, No. 11, Nov. 1976, p. 882.

[2] Barron, A.R., "Predicted squared error: a criterion for automatic model selection," in *Self-Organizing Methods in Modeling, GMDH Type Algorithms*, (S.J. Farlow, Ed.), Marcel Dekker, NY, 1984, pp. 87-103.

[3] Barron, A.R. and R.L. Barron, "Statistical learning networks: a unifying view," *Computing Science and Statistics: Proc. of the 20th Symposium on the Interface*, Am. Statistical Association, (E. Wegman, Ed.), Alexandria, VA, 1988, pp. 192-203.

[4] Barron, R.L., A.N. Mucciardi, F.J. Cook, J.N. Craig, and A.R. Barron, "Adaptive learning networks: development and application in the United States of algorithms related to GMDH," in *Self-Organizing Methods in Modeling. GMDH Type Algorithms*, (S.J. Farlow, Ed.), Marcel Dekker, NY, 1984, pp. 25-65.

[5] Barron, R.L., R.L. Cellucci, P.R. Jordan, III, N.E. Beam, P. Hess, and A.R. Barron, "Applications of polynomial neural networks to FDIE and reconfigurable flight control," *Proc. 42nd Ann. NAECON*, May 1990.

[6] Baxt, W.G., *Trauma: The First Hour*, 1985. (Appleton-Centurycroft, Norwalk, CT)

[7] Baxt, W.G., C.C. Berry, M.D. Epperson, and V. Scalzitti, "The failure of pre-hospital trauma prediction rules to classify trauma patients accurately," *Annals of Emergency Medicine*, Vol. 18, No. 1, Jan. 1989, p. 21.

[8] Baxt, W.G., G. Jones, and D. Fortlage, "The trauma triage rule: a new, resource-based approach to the pre-hospital identification of major trauma victims," *Annals of Emergency Medicine*, Vol. 19, No. 12, Dec. 1990, p. 1401.

[9] Bellamy, R.F., P.A. Maningas, and J.S. Vayer, "Epidemiology of trauma: military experience," *Annals of Emergency Medicine*, Vol. 15. No. 12, Dec. 1986, p. 1384.

[10] Berman, D.A., S.T. Coleridge, and T.A. McMurry, "Computerized algorithm-directed triage in the emergency department," *Annals of Emergency Medicine*, Vol. 18, No. 2, Feb. 1989, p. 141.

[11] Boyd, C.R., M.A. Tolson, and W.S. Copes, "Evaluating trauma care: the TRISS method," *J. Trauma*, Vol. 27, No. 4, Apr. 1987, p. 370.

[12] Bruttig, Stephen J., private communication.

[13] Burke, J.F., R.J. Boyd, and C.J. McCabe, *Trauma Management: Early Management of Visceral, Nervous System, and Musculoskeletal Injuries*, 1988. (Yearbook Medical Publishers, Inc., Boca Raton, FL)

[14] Champion, H.R., W.J. Sacco, R.L. Lepper, E.M. Atzinger, W.S. Copes, and R.H. Prall, "An anatomic index of injury severity," *J. Trauma*, Vol. 20, No. 3, Mar. 1980, p. 197.

[15] Champion, H.R., W.J. Sacco, D.S. Hannan, R.L. Lepper, E.S. Atzinger, W.S. Copes, and R.H. Prall, "Assessment of injury severity: the triage index," *Critical Care Medicine*, Vol. 8, No. 4, Apr. 1980, p. 201.

[16] Champion, H.R., W.J. Sacco, A.J. Carnazzo, W.Copes, and W.J. Fouty, "Trauma score," *Critical Care Medicine*, Vol. 9, No. 9, Sep. 1981, p. 672.

[17] Champion, H.R., "Field triage of trauma patients," *Annals of Emergency Medicine*, Vol. 11, No. 3, Mar. 1982, p. 160.

[18] Champion, H.R., W.J. Sacco, and T.K. Hunt, "Trauma severity scoring to predict Mortality," *World J. Surgery*, Vol. 7, No. 1, Jan. 1983, p. 4.

[19] Champion, H.R., W.J. Sacco, W.S. Copes, D.S. Gann, T.A. Gennarelli, and M.E. Flanagan, "A revision of the trauma score," *J. Trauma*, Vol. 29, No. 5, May 1989, p. 623.

[20] Champion, H.R., W.S. Copes, W.J. Sacco, C.F. Frey, J.W. Holcroft, D.B. Hoyt, and J.A. Weigelt, "Improved predictions from a severity characterization of trauma (ASCOT) over trauma and injury severity score (TRISS): results of an independent evaluation," *J. Trauma*, Vol. 40, No. 1, Jan. 1996, p. 42.

[21] Civil, I.D., and C.W. Schwab, "The abbreviated injury scale, 1985 revision: a condensed chart for clinical use," *J. Trauma*, Vol. 28, No. 1, Jan. 1988, p. 87.

[22] Copes, W.S., H.R. Champion, W.J. Sacco, M.M. Lawnick, S.L. Keast, and L.W. Bain, "The injury severity score revisited," *J. Trauma*, Vol. 28, No. 1, Jan. 1988, p. 69.

[23] Copes, W.S., M.M. Lawnick, H.R. Champion, and W.J. Sacco, "A comparison of abbreviated injury scale 1980 and 1985 Versions," *J. Trauma*, Vol. 28, No. 1, Jan. 1988, p. 78.

[24] Curka, P.A., P.E. Pepe, V.F. Ginger, R.C. Sherrard, M.V. Ivy, and B.S. Zachariah, "Emergency medical services priority dispatch," *Annals of Emergency Medicine*, Vol. 22, No. 11, Nov. 1993, p. 1688.

[25] Duda, R.O., and P.E. Hart, *Pattern Classification and Scene Analyses*, 1973. (John Wiley & Sons: New York)

[26] Emerman, C.L., B. Shade, and J. Kubincanek, "A comparison of EMT judgment and pre-hospital trauma triage instruments," *J. Trauma*, Vol. 31, No. 10, Oct. 1991, p. 1369.

[27] Falk, J.E., S.W. Palocsay, W.J. Sacco, W.S. Copes, and H.R. Champion, "Bounds on a trauma outcome function via optimization," *Operations Research*, Vol. 40, No. 1, Jan.-Feb. 1992, p. S86.

[28] Friedman, J.H. and W. Stuetzle, "Projection pursuit regression," *J. Amer. Statistical Association*, Vol. 76, 1981, pp. 817-823.

[29] Friedman, J.H., "Multivariate adaptive regression splines," *Annals of Statistics*, Vol. 19, No. 1, 1991, pp. 1-66.

[30] Fries, G.R., G. McCalla, M.A. Levitt, and R. Cordova, "A prospective comparison of paramedic judgment and the trauma triage rule in the pre-hospital setting," *Annals of Emergency Medicine*, Vol. 24, No. 5, Nov. 1994, p. 885.

[31] Gabor, D., "Communication theory and cybernetics," *Trans. of IRE*, Vol. CT-1, No. 4, 1954, p. 19.

[32] Gabor, D., P.L. Wilby, and R. Woodcock, "A universal non-linear filter, predictor and simulator which optimizes itself by a learning process," *J. IEE*, paper received Oct. 17, 1959.

[33] Garber, B.G., P.C. Hébert, G. Wells, and J. Yelle, "Validation of trauma and injury severity score in blunt trauma patients by using a Canadian trauma registry," *J. Trauma*, Vol. 40, No. 5, May 1996, p. 773.

[34] Gormican, S.P.,"CRAMS Scale: field triage of trauma victims," *Annals of Emergency Medicine*, Vol. 11, No. 3, Mar. 1982, p. 132.

[35] Guyton, A.C., *Textbook of Medical Physiology*, 1976. (W.B. Saunders, Philadelphia)

[36] Hannan, E.L., J. Mendeloff, L.S. Farrell, G. Cayten, and J.G. Murphy, "Validation of TRISS and ASCOT using a Non-MTOS Trauma registry," *J. Trauma*, Vol. 38, No. 1, Jan. 1995, p. 83.

[37] Hecht-Nielsen, R., *Neurocomputing*, 1990. (Addison Wesley: Reading, MA)

[38] Hedges, J.R., W.J. Sacco, and H.R. Champion, "An analysis of pre-hospital care of blunt trauma," *J. Trauma*, Vol. 22, No. 12, Dec. 1982, p. 989.

[39] Hedges, J.R., S. Feero, B. Moore, D.W. Haver, and B. Shultz, "Comparison of pre-hospital trauma triage instruments in a semirural population," *J. Emergency Medicine*, Vol. 5, 1987, p. 197.

[40] Henderson, J.V., R.K. Pruett, A.R. Galper, and W.S. Copes, "Interactive videodisc to teach combat trauma life support," *J. Medical Systems*. Vol. 10, No. 3, 1986, p. 271.

[41] Hosmer, D.W., and S. Lemeshow, "*Applied Logistic Regression*," 1989. (John Wiley & Sons: Reading, MA)

[42] Hudgins, A.,"Field triage," in *Decision-Making in Trauma Management*, (M.E. Mancini, Ed.), B.C. Decker, Inc., Philadelphia, 1991, p. 16.

[43] Hurst, J.M., "Trauma: an overview," in *Intensive Care Medicine*, 2nd ed., (J.M. Rippe, Ed.), Little Brown & Co., Boston, 1991, p. 1455.

[44] Ivakhnenko, A.G., "The group method of data handling – A rival of stochastic approximation," *Soviet Automatic Control*, Vol. 1, 1968, pp. 43-55.

[45] Ivakhnenko, A.G., "Polynomial theory of complex systems," *IEEE Trans. Systems, Man, & Cybernetics*, Vol. SMC-1, No. 4, Oct. 1971, pp. 364-378.

[46] Jagoda, A., M. Pietrzak, S. Hazen, and J.S. Vayer, "Pre-hospital care and the military," *Military Medicine*, Vol. 157, No. 1, Jan. 1992, p. 11.

[47] Knudson, P., C.A. Frecceri, and S.A. DeLateur, "Improving the field triage of major trauma victims," *J. Trauma*, Vol. 28, No. 5, 1988, p. 602.

[48] Koehler, J.J., L.J. Baer, S.A. Malaga, N.R. Navitskas, and J.E. Huizenga, "Pre-hospital index: a scoring system for field triage of trauma victims," *Annals of Emergency Medicine*, Vol. 15, No. 2, Feb. 1986, p.178.

[49] Larson, E.C., N.A. Nigro, and B.E. Parker, Jr., *Speeding Simulations Via Recurrent Neural Networks*, Barron Associates, Inc. Interim Technical Rept. for Naval Surface Warfare Center, Carderock Division, Contract N00167-94-C-0082, Nov. 1995.

[50] Lemeshow, S., and D.W. Homer, Jr., "A review of goodness of fit statistics for use in the development of logistic regression models," *Am. J. of Epidemiology*, Vol. 115, No. 1, 1982, p. 92.

[51] Lemeshow, S., D. Teres, J.S. Avrunin, and H. Pastides, "Predicting the outcome of intensive care unit patients," *J. Am. Statistical Association*, Vol. 83, No. 402, Jun. 1988, p. 348.

[52] Lemeshow, S., D. Teres, J. Klar, J.S. Avrunin, S.H. Gehlbach, and J. Rapoport, "Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients," *J. Am. Medical Association*, Vol. 270, No. 20, Nov. 24, 1993, p. 2478.

[53] Le Gall, R., S. Lemeshow, and F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," *J. Am. Medical Association*, Vol. 270, No. 24, Dec. 22/29, 1993, p. 2957.

[54] Lett, R.R., J.A. Hanley, and J.S. Smith, "The comparison of injury severity instrument performance using likelihood ratio and ROC curve analyses," *J. Trauma*, Vol. 38, No. 1, Jan. 1995, p. 142.

[55] McGonigal, M.D., J. Cole, C.W. Schwab, D.R. Kauder, M.F. Rotondo, and P.B. Angood, "A new approach to probability of survival scoring for trauma quality assurance," *J. Trauma*, Vol. 34, No. 6, Jun. 1993, p. 863.

[56] MacKenzie, E.J., D.M. Steinwachs, and B. Shankar, "Classifying trauma severity based on hospital discharge diagnoses," *Medical Care*, Vol. 27, No. 4, Apr. 1989, p. 412

[57] Markle, J., C.G. Cayten, D.W. Byrne, F. Moy, and J.G. Murphy, "Comparison between TRISS and ASCOT methods in controlling for injury severity," *J. Trauma*, Vol. 33, No. 2, Aug. 1992, p. 326.

[58] Meredith, W., R. Rutledge, A.R. Hansen, D.W. Oller, M. Thomason, P. Cunningham, and C.C. Baker, "Field triage of trauma patients based upon the ability to follow commands: a study in 29,573 injured patients," *J. Trauma*, Vol. 38, No. 1, 1995, p. 129

[59] Meyer, A.A., W.J. Messick, P.Young, C.C. Baker, S. Fakhry, F. Muakkassa, E.J. Rutherford, L.M. Napolitano, and R. Rutledge, "Prospective comparison of clinical judgment and APACHE II score predicting the outcome in critically ill surgical patients," *J. Trauma*, Vol. 32, No. 6, Jun. 1992, p. 747.

[60] Morris, J.A., P.S. Auerbach, G.A. Marshall, R.F. Bluth, L.G. Johnson, D.D. Trunkey, "The trauma score as a triage tool in the pre-hospital setting," *J. Am. Medical Association*, Vol. 256, No. 10, Sep. 1986, p. 1319.

[61] Ohno-Machado L., "Identification of low frequency patterns in backpropagation neural networks," *Annual Symposium on Computer Applications in Medical Care*, Vol. 9, 1994, p. 853.

[62] Ornato, J.P., E.J. Mlinek, Jr., E.J. Craren, and N. Nelson, "Ineffectiveness of the trauma score and the CRAMS scale for accurately triaging patients to trauma centers," *Annals of Emergency Medicine*, Vol. 14, No. 11, Nov. 1985, p. 1601.

[63] Pearl, J., "Fusion, Propagation, and Structuring in Belief Networks," *Artificial Intelligence*, Vol. 29, 1986, p. 241.

[64] Robertson, C., and A.D. Redmond, *Management of Major Trauma*, 1994. (Oxford Univ. Press, 2nd ed., New York)

[65] Rund, D.A., and T.S. Rausch, *Triage*, 1981. (C.V. Mosby Co., St. Louis)

[66] Rutledge, R., S. Fakhry, E. Rutherford, F. Muakkassa, and A. Meyer, "Comparison of APACHE II, trauma score, and injury severity score as predictors of outcome in critically injured trauma patients," *Am. J. Surgery*, Vol. 166, Sept. 1993, p. 244.

[67] Rutledge, R., "Injury severity and probability of survival assessment in trauma patients using a predictive hierarchical network model derived from ICD-9 codes," *J. Trauma*, Vol. 38, No. 4, 1995

[68] Sacco, W.J., H.R. Champion, P.S. Gainer, S.A. Morelli, S. Fallen, and M.A. Lawnick, "The trauma score as applied to penetrating trauma," *Annals of Emergency Medicine*, Vol. 13, No. 6, Jun. 1984, p. 415.

[69] Sacco, W.J., J.W. Jameson, W.S. Copes, and H.R. Champion, "Partition: a quantitative method for evaluating pre-hospital services for trauma patients," *Comput. Biol. Med.* , Vol. 18, No. 3, Jul. 1988, p. 221.

[70] *Logistic Regression Examples Using the SAS System*, Ver. 6, 1st ed., SAS Institute, Inc., Cary, NC, 1995.

[71] Schaible, S., and T. Ibaraki, "Fractional programming," *Eur. J. Operational Research*, Vol. 12, 1983, p. 325.

[72] Smith, J.S., and M.J. Bartholomew, "Trauma index revisited: a better triage tool," *Critical Care Medicine*, Vol. 18, No. 2, Feb. 1990, p. 174.

[73] Stair, T.O., and J.M. Howell, "How long will it take? How much will it cost?: Multiple regression and neural network programs at ED triage," *Am. J. Emergency Medicine*, Vol. 13, No. 1, Jan. 1995, p. 118.

[74] Summers, S.R., and B. Summers, "Beat the clock," *Emergency Medical Services*, Sept. 1992, p. 27.

[75] Stochetti, N., A. Furlan, and F. Volta, "Hypoxemia and arterial hypotension at the accident scene in head injury," *J. Trauma*, Vol. 40, No. 5, May 1996, p. 764.

[76] Turner, O., R. Rutledge, J. Deis, and E. Bedrick, "ICISS: An international classification of diseases-9 based injury severity score," *J. Trauma*, Vol. 41, No. 3, Sept. 1996, pp. 380-388.

[77] Trunkey, D.D., J. Siegel, S.P. Baker, T.A. Gennarelli, "Panel: current status of trauma severity indices," *J. Trauma*, Vol. 23, No. 3, Mar. 1983, p. 185.

[78] Valenzuela, T.D., "What is 'major trauma?'," *Annals of Emergency Medicine* , Vol. 19, No. 12, Dec. 1990, p. 1470.

[79] Vayer, J.S., R.P. Ten Eyk, and M.L. Cowan, "New concepts in triage" *Annals of Emergency Medicine*, Vol. 15, No. 8, Aug. 1986, p. 927.

[80] Ward, D.G., B.E. Parker, Jr., and R.L. Barron, *Active Control of Complex Systems Via Dynamic (Recurrent) Neural Networks*, Barron Associates, Inc. Task Final Technical Rept. for Office of Naval Research, Contract N00014-89-C-0137, May 1992.

[81] Ward, D.G., B.E. Parker, Jr., and R.L. Barron, "Principles of function estimation using artificial neural networks," Barron Associates, Inc. Technical Memorandum, Sept. 1993.

# A Conventional Scoring Systems

## A.1 Injury Severity Score

The Injury Severity Score (ISS) provides a summary index of the overall severity of injury in terms of the Abbreviated Injury Scale (AIS) scores specific to six anatomical regions:

1 Head/Neck
2 Face
3 Chest
4 Abdomen/Pelvic Contents
5 Extremities/Pelvic Girdle
6 Skin

The severity of injury in each region is coded as follows:

0 None
1 Minor
2 Moderate
3 Serious, but not life-threatening
4 Life-threatening, but survival probable
5 Survival uncertain
6 Survival chances very dim

ISS is defined as the sum of the squares of the three highest AIS scores, provided that all are '5' or less. The highest ISS is therefore $5^2 + 5^2 + 5^2 = 75$. If any one of the AIS scores is '6', an ISS of 75 is assigned automatically.

## A.2 Glasgow Coma Scale

The Glasgow Coma Scale (GCS) provides a summary assessment of neurological condition. Points for eye, verbal, and motor response are summed to obtain the total GCS score:

| | | |
|---|---|---|
| Eye Opening | None | 1 |
| | In response to pain | 2 |
| | In response to voice | 3 |
| | Spontaneous and voluntary | 4 |
| Speech | None | 1 |
| | Incomprehensible | 2 |
| | Inappropriate words | 3 |
| | Confused | 4 |
| | Oriented and alert | 5 |
| Motor | None | 1 |
| | Extension under pain | 2 |
| | Flexion under pain | 3 |
| | Withdrawal from pain | 4 |
| | Purposeful movement under pain | 5 |
| | Voluntarily obeys commands | 6 |

| | | |
|---|---|---|
| Total GCS | | 3 – 15 |

## A.3 Original Champion Trauma Score (TS)

| | | |
|---|---|---|
| RR | ≥ 36* | 2 |
| | 25 – 35 | 3 |
| | 10 – 24 | 4 |
| | 1 – 9 | 1 |
| | None | 0 |
| Respiratory Effort | Normal | 1 |
| | Shallow or retractive | 0 |
| SBP | ≥ 90† | 4 |
| | 70 – 89 | 3 |
| | 50 – 69 | 2 |
| | 1 – 49 | 1 |
| | 0 | 0 |
| Capillary return | Normal | 2 |
| | Delayed | 1 |
| | None | 0 |
| GCS | 14 – 15 | 5 |
| | 11 – 13 | 4 |
| | 8 – 10 | 3 |
| | 5 – 7 | 2 |
| | 3 – 4 | 1 |
| Total TS | | 1 – 16 |

## A.4 Baxt Trauma Triage Rule (TTR)

The Baxt TTR score is defined as 1 if any of the following criteria are met, 0 otherwise:

- SBP < 85 mm Hg
- Glasgow Motor Score < 5
- Penetrating cranial, neck, or thoracic injury

## A.5 CRAMS

| | | |
|---|---|---|
| Circulation | Normal capillary refill and SBP > 100 | 2 |

*RR measured breaths per minute

†SBP measured in millimeters of mercury

| | | |
|---|---|---|
| | Delayed capillary refill and $85 < SBP < 100$ | 1 |
| | No capillary refill and $SBP < 85$ | 0 |
| Respiration | Normal | 2 |
| | Labored or shallow | 1 |
| | None | 0 |
| Abdomen | Abdomen and thorax nontender | 2 |
| | Abdomen or thorax tender | 1 |
| | Abdomen rigid or chest flail | 0 |
| Motor | Normal | 2 |
| | Response to pain only | 1 |
| | No response or decerebrate | 0 |
| Speech | Normal | 2 |
| | Confused | 1 |
| | None or garbled | 0 |
| CRAMS score | | $0 - 10$ |

## A.6 Pre-Hospital Index (PHI)

| | | |
|---|---|---|
| SBP | $> 100$ | 0 |
| | $86 - 100$ | 1 |
| | $75 - 85$ | 2 |
| | $< 75$ | 5 |
| Pulse | $\geq 120^{\ddagger}$ | 3 |
| | $51 - 119$ | 0 |
| | $< 50$ | 5 |
| Respirations | Normal | 0 |
| | Labored or shallow | 3 |
| | $< 10$/min or needs intubation | 5 |
| Consciousness | Normal | 0 |
| | Confused or combatitive | 3 |
| | None | 5 |
| Penetrating abdominal or thoracic injury | Yes | 4 |
| | No | 0 |
| PHI score | | $0 - 24$ |

[‡]Pulse (synonymous with heart rate) measured in beats per minute

## A.7 Revised Trauma Index (RTI)

| | | |
|---|---|---|
| Injured region | Limbs or skin | 1 |
| | Back | 3 |
| | Chest | 5 |
| | Head, abdomen, or multiple injuries | 6 |
| Type of wound | Minor open wound | 1 |
| | Single blunt impact or second-degree burn | 3 |
| | Major open wound, stab wound, third-degree burn | 5 |
| | Gunshot wound or multiple blunt impacts | 6 |
| SBP and Pulse | SBP $>$ 100 and pulse $<$ 100 | 1 |
| | SBP80 $-$ 100 and pulse100 $-$ 140 | 3 |
| | SBP $<$ 80 and pulse $>$ 140 | 5 |
| | no pulse | 6 |
| RR | $10 - 24$ | 1 |
| | $25 - 35$ | 3 |
| | $>$ 35 or $<$ 10 | 5 |
| | apnea | 6 |
| Consciousness | Drowsy, disoriented, or confused | 1 |
| | responsive to voice | 3 |
| | responsive to pain only | 5 |
| | unresponsive | 6 |
| RTI score | | $4 - 30$ |

## A.8 RR/Pulse/Motor Score (RPM)

| | | |
|---|---|---|
| RR | same as in TS | $1 - 6$ |
| Pulse | $\geq 120$ | 3 |
| | $61 - 120$ | 4 |
| | $41 - 60$ | 2 |
| | $1 - 40$ | 1 |
| | None | 0 |
| Motor | Obeys command | 4 |
| | Responsive to pain | 3 |
| | Withdrawal from pain | 2 |
| | Flexion or extension under pain | 1 |
| | None | 0 |

| | | |
|---|---|---|
| RSM score | | $0-12$ |

## A.9 RR/SBP/Motor Score (RSM)

| | | |
|---|---|---|
| RR | same as in TS | $0-4$ |
| SBP | same as in TS | $0-4$ |
| Motor | same as in RPM | $0-4$ |
| RSM score | | $0-12$ |

## A.10 RR/SBP/GCS Score (RSG)

| | | |
|---|---|---|
| RR | $10-29$ | 0 |
| | else | 1 |
| SBP | $\geq 90$ | 0 |
| | else | 1 |
| GCS | $14-15$ | 0 |
| | else | 1 |
| RSG score | | $0-3$ |

## A.11 Mechanism of Injury (MOI)

MOI = 1 if any of the following criteria are met, 0 otherwise:

- Extrication time from vehicle
- Vehicle occupant forcefully thrown
- Pedestrian thrown by impact with motor vehicle
- Fall of greater than 15 feet
- Penetrating wound (excluding extremities)

## A.12 Gestalt Impression of Severity as Estimated by Paramedic (SEV)

The SEV score ranges from 1 to 3, based on the following impressions:

- 1 = Not serious
- 2 = Potentially life-threatening
- 3 = Critically life-threatening

## A.13 Kane's Revised Checklist (KRC)

KRC = 1 if any of the following criteria are met, 0 otherwise:

- No spontaneous, voluntary eye opening
- Abnormal capillary refill
- Penetrating cranial, neck, or thoracic injury

## A.14 Revised Trauma Score (RTS and T-RTS)

| | | |
|---|---|---|
| RR | $\geq 30$ | 3 |
| | $10-29$ | 4 |
| | $6-9$ | 2 |
| | $1-5$ | 1 |
| | None | 0 |
| SBP | $\geq 90$ | 4 |
| | $76-89$ | 3 |
| | $50-75$ | 2 |
| | $1-49$ | 1 |
| | 0 | 0 |
| GCS | $13-15$ | 4 |
| | $9-12$ | 3 |
| | $6-8$ | 2 |
| | $4-5$ | 1 |
| | 3 | 0 |
| T-RTS Score | | $0-12$ |

The Revised Trauma Score for triage (T-RTS) is the sum of the three coded values of GCS, RR, and SBP. Using MTOS data, Champion [39] performed a logistic regression to relate probability of survival, $P_s$, to these coded variables. Their curve fit yielded

$$P_s = \frac{1}{1+e^{-L}}$$

in which $L = -3.5718 + 0.9368 \times \mathrm{GCS}_c + 0.7326 \times \mathrm{SBP}_c + 0.2908 \times \mathrm{RR}_c$. The subscript denotes coded values. The Revised Trauma Score (RTS) is defined as the logit polynomial, L, less the

constant term, *viz.*,

$$\mathrm{RTS} = 0.9368 \times \mathrm{GCS}_c + 0.7326 \times \mathrm{SBP}_c + 0.2908 \times \mathrm{RR}_c$$

## A.15 Trauma and Injury Severity Score (TRISS)

TRISS computes a probability of survival, $P_s$ via the logistic formula

$$P_s = \frac{1}{1 + e^{-L}}$$

in which $L = c_0 + c_1 \times \mathrm{RTS} + c_2 \times \mathrm{ISS} + c_3 \times \mathrm{AGE}$, where $\mathrm{AGE}_c = 1$ if $\geq 55$ years, $\mathrm{AGE}_c = 0$ otherwise. On MTOS data, different sets of coefficients were fitted separately for blunt and penetrating injuries:

| | Blunt | Penetrating |
|---|---|---|
| $c_0$ | −1.2470 | −0.6024 |
| $c_1$ | 0.9544 | 1.1430 |
| $c_2$ | −0.0768 | −0.1516 |
| $c_3$ | −1.9052 | −2.6676 |

Note that since TRISS incorporates ISS, it *cannot* be used as a pre-hospital triage tool. Instead, it is normally used for *ex post* quality-of-care evaluation.

## A.16 Severity Characterization of Trauma (ASCOT)

As in TRISS, probability of survival is computed from a logistic formula, with coefficients fitted to MTOS data, *viz.*,

$$P_s = \frac{1}{1 + e^{-L}}$$

in which

$$L = c_0 + c_1 \times \mathrm{GCS}_c + c_2 \times \mathrm{SBP}_c + c_3 \times \mathrm{RR}_c + c_4 \times \mathrm{A} + c_5 \times \mathrm{B} + c_6 \times \mathrm{C} + c_7 \times \mathrm{AGE}$$

The coding convention is the same as in TRISS. The variables A, B, and C refer to the severities of anatomical injuries in specific regions:

- A Cranial or spinal cord injury of severity in AIS range $3 - 5$
- B Thoracic or frontal neck injury in AIS range $3 - 5$
- C Any other injury in AIS range $3 - 5$

A, B, or C is equal to 1 if the corresponding criterion is met, 0 otherwise. As in TRISS, ASCOT uses separate sets of coefficients for blunt and penetrating cases:

| | Blunt | Penetrating |
|---|---|---|
| $c_0$ | −1.1570 | −1.1350 |
| $c_1$ | 0.7705 | 1.0626 |
| $c_2$ | 0.6583 | 0.3638 |
| $c_3$ | 0.2810 | 0.3332 |
| $c_4$ | −0.3002 | −0.3702 |
| $c_5$ | −0.1961 | −0.2053 |
| $c_6$ | −0.2086 | −0.3188 |
| $c_7$ | −0.6355 | −0.8365 |

## A.17 Proposed Triage Rules

The following decisional rules have been proposed in the literature [39] as tools for pre-hospital triage. Reported sensitivity and specificity performances are as follows:

| Send to Trauma Center if: | Sensitivity (%) | Specificity (%) |
|---|---|---|
| TS $\leq 14$ | 63 | 88 |
| TS $\leq 12$ | 46 | 97 |
| MOI $= 1$ | 54 | 93 |
| TS $\leq 12$ or GCS $\leq 10$ or MOI $= 1$ | 78 | 63 |
| SEV $= 3$ | 51 | 96 |
| CRAMS $\leq 8$ | 39 | 89 |
| PHI $\geq 4$ | 73 | 75 |
| RSM $\leq 10$ | 59 | 92 |
| RPM $\leq 10$ | 61 | 88 |
| KRC $= 1$ | 85 | 65 |
| T-RTS $\leq 11$ | 59 | 82 |
| T-RTS $\leq 10$ | 49 | 92 |
| T-RTS $\leq 9$ | 39 | 96 |
| RTI $\geq 15$ | 95 | 87 |
| TTR $= 1$ | 92 | 92 |

# B Trauma Physiology Examples

In this appendix, we discuss a few representative examples [6] of the physiological implications of trauma. As a primary illustration, let us consider the consequences and manifestations of *hypovolemia*, i.e., significant loss of blood volume due to exsanguination (bleeding). Physiological responses at both the macro- and micro-circulatory levels take place. The sympathetic nervous system intervenes to induce vasoconstriction of the arterioles and to reduce the storage capacitance of the veins. Sympathetic stimulation of cardiac muscle increases the heart rate, respiratory rate, and myocardial contractility to offset the diminution of stroke volume and systolic blood pressure. Blood flow is allocated preferentially to those vital organs, namely the heart, lungs, and brain, least tolerant of oxygen debt. At the microvascular level, constriction of the arterioles diminishes the hemostatic pressure within the capillaries, perturbing the osmotic equilibrium and inducing movement of water from the interstitial spaces into the vessels. This effect is manifested by dilution of the hematocrit (volumetric fraction of blood comprising red cells) and the serum proteins. In this way, the body is able to establish a state of *compensated shock* against moderate (10-15%) losses of blood volume. Through frequent fluid intake and endogeneous release of aldosterone and antidiuretic hormone (which modulate the fluid-removal function of the kidneys), the patient is able to hold out for a relatively long period, without intensive medical attention, until a compatible blood donation can be found to restore normal volume.

If exsanguination persists, however, the compensation mechanisms are eventually overridden. Patients in such states of progressive shock exhibit elevated levels of catecholamines and kinins (vasoactive agents that increase capillary permeability and reduce venous capacitance to facilitate release of blood stored in the veins). At the cellular level, tissues eventually become *ischemic*, i.e., unable to receive the supply of nutrients necessary for normal metabolism. The cells resort to anaerobic metabolism, which results in diminished ATP (adenosine triphosphate) production, hypercarbia, and lactic acidosis. The most ominous consequence is that the ATP deficit deprives the cells of their ability to remove sodium, creating an osmotic gradient that favors fluid movement into the cells from the vessels. This counteracts the intravascular fluid movement established in the compensated state and has the effect of exacerbating the hypovolemia. Patients in this condition have usually reached an unsalvageable state of irreversible shock.

Hypovolemia thus illustrates the complicated and dynamically rich character of the response of the body to trauma. Interpreted as dynamical system behavior, the physiological defense activity is highly nonlinear with respect to severity of blood loss. For example, the compensated state may be regarded as a metastable dynamical state that can be maintained with little medical intervention (i.e., fluid intake and direct pressure to prevent further bleeding). There is a critical severity of hypovolemia, however, which becomes fatal (namely where the ionic pump mechanisms fail). Cardiogenic failure, where the myocardium can no longer supply adequate cardiac output, and renal failure, where the kidneys themselves become ischemic, pose additional potential threats to life in hypovolemic cases. The analysis also demonstrates how various biological indicators (tachycardia, tachypnea, reduced pulse pressure, lactic acidosis, depressed arterial $PO_2$, elevated venous $PCO_2$, low hematocrit, high catecholamine levels) can furnish telltale signs of the abnormal physiological condition characterizing the patient.

Note that all of these outputs are rather difficult to measure and may require invasive procedures. Just looking to see whether the patient has bled through an open wound provides little help in the numerous cases where patients have sustained blunt, but nonetheless life-threatening, trauma.

Two examples in point are *hematomas*, or large internal blood balloons, stemming respectively from blunt head and thoracic injuries. Pulmonary hematomas can have at least two deadly effects, namely *tension pneumothorax*, in which inspired air accumulates in the chest cavity but cannot escape, and *cardiac tamponade*, in which the hematoma mass presses against the heart, constricting diastolic filling and forcing heart rate to rise. In tension pneumothorax, a performation in lung tissue causes inspired air to become trapped in a pocket just outside the lung. The accumulation of air within the pocket, which increases with each breath, constricts the effective volume of the lung and results in death within minutes unless the pocket is punctured from without to release the air. Both conditions require astute perception to be recognized and fast reaction to avert death.

In closed-head trauma, a different host of pathophysiological effects arises. Cerebral hemorrhaging raises *intracranial pressure* (ICP), which impairs venous outflow, blocks egress of cervospinal fluid, and causes *cerebral perfusion pressure* (CPP, defined as systemic arterial pressure less ICP) to fall. Equally importantly, mental consciousness diminishes, and the patient tends to fall into a hypoventilatory state known as *syncope*. Hypoxia, hypercarbia, and metabolic acidosis gradually set in. As a compensatory phenomenon somewhat analogous to compensated hypovolemic shock, the *Cushing mechanism* attempts to maintain CPP by raising systemic arterial pressure (averaged over both systole and diastole). If this mechanism is overwhelmed, however, cerebral blood flow is impaired, and ischemic neuronal cells perish rapidly. For head injuries, therefore, many of the same indicators useful in the cases of hypovolemia and and pulmonary hematomas (e.g., SBP, DBP, RR, P, CO, $PaO_2$, $PaCO_2$, tissue pH) provide pertinent information, along with several others specific to head injury (i.e., ICP, CPP, and level of consciousness, as measured by EY, VB, and MT).

# C Theory of Classification Models

## C.1 Class Membership Probabilities

As a first-principles starting point for introducing the mathematical theory of classification, let us consider a simple case in which there exist two classes (POSITIVE and NEGATIVE) and a single independent attribute variable, $x$, which is real-valued and continuous. Let us assume that the *probability densities* describing the distributions of the two classes with respect to $x$ are Gaussian, *viz.*,

$$P_{\rm n}(x) = \sigma^{-1}(2\pi)^{-1/2}\exp\left[-(x-\mu_{\rm n})^2/2\sigma^2\right] \quad (7a)$$

$$P_{\rm p}(x) = \sigma^{-1}(2\pi)^{-1/2}\exp\left[-(x-\mu_{\rm p})^2/2\sigma^2\right]. \quad (7b)$$

Eq. 7a states that for a randomly selected observation belonging to the NEGATIVE class, the probability that its attribute value lies between $x$ and $x+\Delta x$ is approximately $P_{\rm n}(x)\,\Delta x$ if $\Delta x$ is sufficiently small. The NEGATIVE and POSITIVE probability densities are both Gaussian centered at $\mu_{\rm n}$ and $\mu_{\rm p}$ respectively. For reasons that will soon become apparent, the variances for the two are assumed to be the same. The distributions are illustrated generically in Fig. 9, from which overlap is apparent ($\mu_{\rm n}=0$, $\mu_{\rm p}=3$, $\sigma=1$). The ease with which the classes can be distinguished depends on the separation of the peak centers, $|\mu_{\rm p}-\mu_{\rm n}|$, relative to the variance, $\sigma^2$.

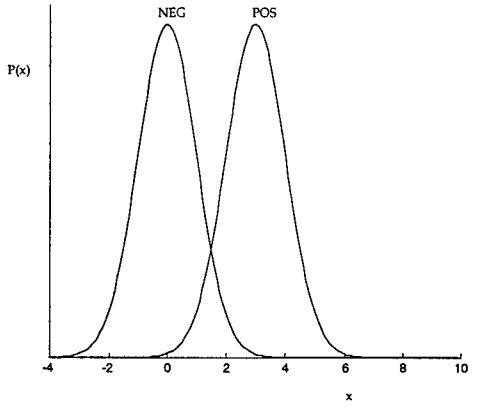


**Figure 9: Class Probability Distributions in Attribute Space**

The objective of classification is to determine the class membership probabilities of an arbitrary exemplar, given that its attribute assumes value $x$. These probabilities may be obtained by way of Bayes' formula (see Appendix C.2 below), *viz.*,

$$\pi_{\rm n}(x) = \alpha_{\rm n}P_{\rm n}(x)\left[\alpha_{\rm n}P_{\rm n}(x)+\alpha_{\rm p}P_{\rm p}(x)\right]^{-1} \quad (8a)$$

$$\pi_{\rm p}(x) = \alpha_{\rm p}P_{\rm p}(x)\left[\alpha_{\rm n}P_{\rm n}(x)+\alpha_{\rm p}P_{\rm p}(x)\right]^{-1}. \quad (8b)$$

$\pi_{\rm n}(x)$ is the probability that an arbitrary exemplar of attribute value $x$ belongs to the NEGATIVE class. $\pi_{\rm p}(x)=1-\pi_{\rm n}(x)$ is the probability that that same exemplar belongs to the POSITIVE

class. $\alpha_n$ and $\alpha_p$ are *a priori* probabilities ($\alpha_n + \alpha_p = 1$), i.e., the probabilities of an any arbitrary exemplar belonging to the respective classes. For example, $\alpha_n = 0.5$ and $\alpha_p = 0.5$ means that NEGATIVEs and POSITIVEs are equally prevalent in the general population of interest (e.g., for epidemiological studies).

Substituting Eqs. 7 into Eqs. 8 yields:

$$\pi_n(x) = e^{-\underline{\theta}^T \cdot \underline{x}} \left[1 + e^{-\underline{\theta}^T \cdot \underline{x}}\right]^{-1} \tag{9a}$$

$$\pi_p(x) = \left[1 + e^{-\underline{\theta}^T \cdot \underline{x}}\right]^{-1} \tag{9b}$$

in which $\underline{x} \equiv [1 \quad x]^T$, and $\underline{\theta} = [\theta_1 \quad \theta_x]^T \equiv \left[\ln(\alpha_p/\alpha_n) - (\mu_p^2 - \mu_n^2)/2\sigma^2 \quad (\mu_p - \mu_n)/\sigma^2\right]^T$. The class membership probabilities in Eqs. 9 are *logistic*, or sigmoidal, functional forms in $x$. Assuming, for concreteness, that $\mu_p > \mu_n$, it follows from the logistic forms that $\pi_n \to 1$, $\pi_p \to 0$ as $x \to -\infty$ and $\pi_n \to 0$, $\pi_p \to 1$ as $x \to \infty$. Thus, the POSITIVE class is dominant for $x \gg 0$ and the NEGATIVE class is dominant for $x \ll 0$. Were the variances in Eqs. 7 unequal for the two classes, the class membership probabilities would not assume logistic forms and no such regions of attribute space in which a given class dominates could be established.

From a training database, in which attribute values and actual class membership are provided for each exemplar, the coefficients, $\underline{\theta}$, can be computed either from first-principles analysis (i.e., by determining the probability densities, fitting Gaussian distributions to them, and using the above formula for $\underline{\theta}$ in terms of the peak centers and variances) or via logistic regression, which is the analogue of least-squares regression for classification problems. For a general classification problem involving $C \geq 2$ classes, the class membership probabilities are logistic forms, *viz.,*

$$\pi_c(\underline{x}) = e^{-\underline{\theta}_c^T \cdot \underline{x}} \left[\sum_{c'=1}^{C} e^{-\underline{\theta}_{c'}^T \cdot \underline{x}}\right]^{-1} \tag{10}$$

with $\underline{\theta}_C = \underline{0}$ to guarantee uniqueness of the fitted coefficients. Arguments of logistic functions, such as $\underline{\theta}_c^T$ in Eq. 10, are called *logits*. Having obtained fitted coefficients for the training database, the polynomial functions $\underline{\theta}_c^T \cdot \underline{x}$ for $c = 1, \ldots, C-1$ segregate the classes in attribute space and entirely determine the membership probabilities. Surfaces of constant $\underline{\theta}_c^T \cdot \underline{x}$ serve as natural dividing boundaries that partition the attribute space into regions in which the classes assume various degrees of relative dominance. In the general case, for instance, Class 1 is dominant in regions where $\underline{\theta}_1^T \cdot \underline{x} \ll \underline{\theta}_2^T \cdot \underline{x}, \ldots, \underline{\theta}_C^T \cdot \underline{x}$.

## C.2 Bayes' Theorem

Let us suppose that we are analyzing a population of patients, such as in the example above, whose observable medical condition is characterized by a single attribute variable, $x$, which is continuous and readily measurable. For an arbitrary patient belonging to class $c$ (among $C \geq 2$ possible classes), the probability that the attribute lies between $x$ and $x + \Delta x$ (where $\Delta x$ is sufficiently small) is equal to $P_c(x)\,\Delta x$. $x$ may denote a variable such as respiratory rate, whereas $c$ might denote a triage category.

The expression $P_c(x)\,\Delta x$ is a *conditional probability*, namely the probability that the attribute lies between $x$ and $x + \Delta x$ given that the patient belongs to class $c$. In classical probability theory,

the conditional probability is defined as

$$P_c(x)\,\Delta x \equiv \frac{P(\text{membership in class } c \text{ and attribute lies between } x \text{ and } x+\Delta x)}{P(\text{membership in class } c)}. \tag{11}$$

To compute the probability expressions in the numerator and denominator in Eq. 11, one must envision a large database of patients representing the general population of interest. The ensemble of patients must represent, comprehensively and accurately, the statistical distributions of attributes and class memberships one would expect to encounter in practice, e.g., in modeling the demand distribution in a queueing problem. The denominator, in this analogy, denotes the fraction of patients out of the entire trauma population (e.g., encountered by a given hospital in a particular year) who happen to belong to class $c$, regardless of $x$. This involves treating the database as a long table and counting the percentage of patients in class $c$. In the language of Bayesian probability and belief networks, this is an *a priori* probability, $\alpha_c$.

Given $x$, it will generally be the case that there exist several classes to which the patient could conceivably belong. In Fig. 9, for instance, a value of $x = 2$ could belong to either the POSITIVE or the NEGATIVE class, since the respective probability distributions overlap. Whereas it may be possible to compute $P_c(x)$ from *ex post* epidemiological studies or by other means, the quantity of practical interest in critical care applications is the probability, $\pi_c(x)$, that a patient belongs to class $c$ given that his/her attribute lies sufficiently close to $x$. This is also a conditional probability, but with the circumstances transposed. It is computed as

$$\pi_c(x) = \frac{P(\text{membership in class } c \text{ and attribute lies between } x \text{ and } x+\Delta x)}{P(\text{attribute lies between } x \text{ and } x+\Delta x)}. \tag{12}$$

The numerator in Eq. 12 is the same as that in Eq. 11 and is thus equal to $\alpha_c P_c(x)\Delta x$. The denominator is equal to the probability, with respect to the entire population, that the attribute lies between $x$ and $x + \Delta x$, regardless of class membership. This quantity may be computed by summing the expression in the numerator of Eq. 12 over all classes, *viz.*,

$$P(\text{attribute lies between } x \text{ and } x+\Delta x) = \sum_{c'=1}^{C} \alpha_{c'} P_{c'}(x)\,\Delta x\ . \tag{13}$$

Eq. 12 then becomes

$$\pi_c(x) = \frac{\alpha_c P_c(x)}{\sum_{c'}\ \alpha_{c'} P_{c'}(x)} \tag{14}$$

in which $\Delta x$'s in the numerator and denominator cancel. The expression in Eq. 14 is known as *Bayes' formula* [25], which provides a readymade mechanism for inverting conditional probabilities.

## C.3 Logistic Regression

To illustrate the mechanics of logistic regression by reference to a generic example, let us suppose that we have a training database of $N$ training exemplars of the form $\{\underline{x}_i, y_i\}$, in which $\underline{x}_i$ is a $P \times 1$ set of (synthetic) input variables for the $i$'th exemplar and $y_i$ denotes its actual class membership. Whereas the inputs are observable on line, as in least-squares estimation, class membership can be directly ascertained on line. We wish to develop a model for inferring or predicting the probability,

$\pi_c(\underline{x}_i)$, that the $i$'th observation belongs to class $c$, in which there exist $C \geq 2$ classes, all of which must be represented in the training database.

Logistic regression postulates class membership probability functions of the logistic forms in Eq. 10, with $\underline{\theta}_C = \underline{0}$. The objective is to find coefficients, $\underline{\theta}$, such that the resulting membership probabilities accurately reflect the character of the training exemplars *in toto*. Eq. 10 indicates how to compute the membership probabilities for an arbitrary exemplar given $\underline{x}$ and regression coefficients $\underline{\theta}$ for each class. To fit coefficients to the training database, logistic regression appeals to a *maximum likelihood* principle, in which the negative of the total log-likelihood function of the form

$$\Lambda(\underline{\theta}_1, \ldots, \underline{\theta}_C) \equiv \sum_{i=1}^{N} (y_i = c) \ln \pi_c(\underline{x}_i; \underline{\theta}_1, \ldots, \underline{\theta}_{C-1}) \tag{15}$$

is minimized globally in coefficient space. In Eq. 15, the exponent $(y_i = c)$ is equal to unity (zero) if the statement that the $i$'th observation actually belongs to the class $c$ is true (false). Since each observation belongs to exactly one class, all but one term in Eq. 15 is equal to unity.

To minimize $\Lambda$, its gradient with respect to each $\underline{\theta}$ vector must vanish, *viz.*,

$$(\nabla\Lambda)_{(c,p)} = \partial\Lambda/\partial\theta_{c,p} = \sum_{i=1}^{N} x_{i,p} \left\{ \sum_{c'=1}^{C} \left[ (y_i = c') - \pi_{i,c'} \right] \right\} = 0 \tag{16}$$

for each $p \in \{1, \ldots, P\}$ and $c \in \{1, \ldots, C-1\}$. That $\Lambda$ be a local minimum, not just an extremum, requires that the Hessian tensor be positive definite. The Hessian components compute to

$$(\square\Lambda)_{(c,p),(c',p')} = \partial^2\Lambda/\partial\theta_{c,p}\partial\theta_{c',p'} = \sum_{i=1}^{N} x_{i,p} x_{i,p'} \pi_{i,c'} (\delta_{c,c'} - \pi_{i,c}) \tag{17}$$

in which $\delta_{c,c'} = 1$ if $c = c'$, 0 otherwise. In coefficient space, $\nabla\Lambda$ and $\square\Lambda$ are respectively a column vector and square matrix, both of dimensionality $P(C-1)$.

Eq. 16 is a set of transcendental equations that can be solved only by iterative numerical methods such as Newton-Raphson, in which one solves approximately for the point at which the gradient vanishes by appealing to a first-order Taylor series expansion, *viz.*,

$$(\nabla\Lambda)|_{\underline{\theta}_{j+1}} = (\nabla\Lambda)|_{\underline{\theta}_j} + \Delta\underline{\theta} \cdot (\square\Lambda)|_{\underline{\theta}_j} \tag{18}$$

$\Delta\underline{\theta} \equiv \underline{\theta}_{j+1} - \underline{\theta}_j$ is the difference between the $j$'th and $(j+1)$'th iterative approximations. From Eq. 18, it follows that the gradient in the $(j+1)$'th approximation will be much closer to zero than in the $j$'th if $\Delta\underline{\theta}$ is such that the right-hand side nearly vanishes. This can be accomplished by choosing $\Delta\underline{\theta}$ such that

$$\Delta\underline{\theta} = -(\square\Lambda)^{-1}|_{\underline{\theta}_j} \cdot (\nabla\Lambda)|_{\underline{\theta}_j} \tag{19}$$

in which the dot denotes matrix multiplication. In the Newton-Raphson method, one starts with $\underline{\theta}_0 = 0$ as the zeroth approximation and uses Eq. 19 to obtain successively more accurate approximations of the set of $\underline{\theta}$ values that yield the optimal maximum-likelihood fit. Eqs. 16 and 17 are used to compute the gradient and Hessian at each step.

Convergence is rapid and reliable except in cases involving small training databases where it is possible to completely segregate two or more classes by way of a hyperplane along which $\underline{\theta}_c^T \cdot \underline{x}$ is constant. If in the simple two-class case in Fig. 9, for example, a small training database were such that every single exemplar with $x < 2$ happened to be NEGATIVE and every exemplar with $x \geq 2$ happened to be POSITIVE, the coefficient $\theta_x$ would diverge to infinity. The resulting membership probabilities are still valid, but the logistic regression requires a build-in criterion to break out of the infinite Newton-Raphson loop.

# D Decisional Algorithms and ROC Curves

## D.1 Decisional Algorithms and Classifier Probabilities

Decisional algorithms are coupled intimately with classifiers in that they utilize computed class membership probabilities as the basis for pragmatic decision-making and subsequent action. A decisional algorithm issues a command for the user to take action based on the *working hypothesis* that the exemplar in question (i.e., the patient) belongs to that one class that it selects, or *declares*. For example, if an accident victim has been determined, from a diagnostic classification algorithm, to have a 70% chance of having sustained life-threatening major trauma, but a 30% chance of having only minor injuries, a decisional algorithm would issue the command to treat and evacuate the individual as if he/she actually were a major trauma case. At least for the short run, all eggs are placed in the major trauma basket; efforts would be made to transport the individual to a Level I trauma center. Note, however, that the decision is risky to the extent that it must be based on what the underlying medical condition *suspected* to be as long as the classes are fundamentally difficult to distinguish.

To formulate a decisional rule, one must specify *thresholds* on either the logit polynomials, $\underline{\theta}_c^T \cdot \underline{x}$, in the logistic formulae (Eq. 10) or directly on attribute space. For a particular decisional algorithm tested on a given evaluation database, the *efficacy* of the algorithm, in conjunction with any accompanying classification or estimation algorithms, is summarized by way of the $C \times C$ *confusion matrix*, *viz.*,

$$\underline{\kappa} = \begin{pmatrix} \kappa_{\mathrm{n,n}} & \kappa_{\mathrm{n,p}} \\ \kappa_{\mathrm{p,n}} & \kappa_{\mathrm{p,p}} \end{pmatrix} \tag{1}$$

in which $\kappa_{c,c'}$ denotes the number of actual class $c'$ exemplars assigned to class $c$. Whereas the diagonal elements tally correct decisions, the off-diagonal elements correspond to Type I and Type II errors. The decisional algorithm must be formulated carefully and deliberatively in such a way that the prevalence of Type I and Type II errors are jointly held down to tolerable levels. This is important because penalties for misclassifying cases are typically very severe and must be addressed explicitly. In medicine, this is a matter of life and death; the goal of triage itself is to reduce the prevalence of one type of classification error (overtriage) without increasing appreciably that of the opposite type of misclassification (undertriage). Astute placement of thresholds is necessary.

## D.2 Classification Performance and ROC Curves

Since the confusion matrix elements, in general, scale proportionally with the size of the test database, the efficacy of the decisional algorithm is best revealed by way of two key sets of ratios, the first of which are the elements of the $C \times C$ *classification performance* matrix, $\Pi$, whose generic component $\Pi_{c,c'}$ denotes the percentage of (actual) class $c'$ observations assigned to class $c$ by the decisional algorithm, *viz.*

$$\Pi_{c,c'} = \frac{\kappa_{c,c'}}{\sum_{c''} \kappa_{c'',c'}}. \tag{2}$$

The denominator is the sum of the elements in the $c'$th column of $\underline{\kappa}$. It follows that each column of $\Pi$ sums to unity. The classification error rates, i.e., off-diagonal components of $\Pi$, depend on the choice of thresholds as well as the inherent overlap of the probability densities (as in Fig. 9).

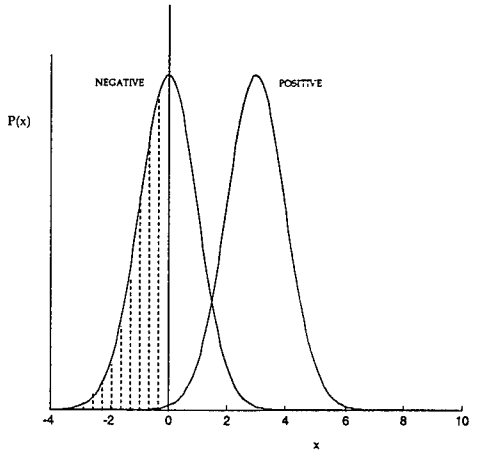


**Figure 10: Probability Distributions for Two-Class Univariate Case with Threshold at $x = 0$**

To illustrate how thresholds are established and classification errors quantified, let us return to the simple univariate two-class problem. Let $\xi$ be a threshold such that the decisional algorithm declares a patient to be NEGATIVE (i.e., not in need of critical care services) if $x < \xi$, and POSITIVE otherwise. Computation of the components of $\Pi$ requires integration under the probability distribution functions. For example, $\Pi_{n,n}$ is the definite integral of $P_n(x)$ from $-\infty$ to $\xi$, i.e., the shaded region in Fig. 10. The matrix computes to

$$\Pi = \begin{pmatrix} \Pi_{n,n} & \Pi_{n,p} \\ \Pi_{p,n} & \Pi_{p,p} \end{pmatrix} = \begin{pmatrix} N[(\xi - \mu_n)/\sigma] & N[(\xi - \mu_p)/\sigma] \\ 1 - N[(\xi - \mu_n)/\sigma] & 1 - N[(\xi - \mu_p)/\sigma] \end{pmatrix} \tag{3}$$

in which

$$N(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du \tag{4}$$

is the definite integral of the normalized unit-variance Gaussian. To obtain good classification (large diagonal and small off-diagonal matrix elements), it is clearly necessary to place the threshold between the two peaks. Classification results are customarily presented using *receiver operating characteristic* (ROC) curves, in which the *sensitivity*, $\Pi_{p,p}$, is plotted against the *specificity*, $\Pi_{n,n}$. Sensitivity and specificity respectively are the fractions of actual POSITIVEs and NEGATIVEs correctly identified as such; for this reason, they are extremely important for *ex post* quality control and evaluation of algorithm performance.

Fig. 11 illustrates a family of nominal ROC curves, which differ in the ratio of $\Delta\mu \equiv \mu_p - \mu_n$ to $\sigma$. In the plot, nine curves are shown with $\Delta\mu/\sigma$ ranging from 0 to 2.0 in steps of 0.25. For cases in which $\Delta\mu/\sigma$ is large, the ROC curve is tightly wedged into the upper left-hand corner of the plot square. The two classes are easily distinguishable, and it is possible to achieve high sensitivity and specificity simultaneously. The opposite extreme is the 45° line from the lower left to the upper right, in which case the peak centers coincide and the classes are completely indistinguishable. Different points along a single ROC curve reflect the classification performance accruing to various threshold placements. The upper right-hand corner corresponds to $\xi = -\infty$, in which case all observations are declared POSITIVE. At this conservative end of the threshold

spectrum, high sensitivity at the expense of low specificity results in overtriage. The lower left-hand corner corresponds to $\xi = \infty$, which is the undertriage extreme. The midpoints of the curves, lying on the downsloping 45° line from the upper left to the lower right, correspond to thresholds placed exactly halfway between the peak centers; the ROC curves are all symmetric about this diagonal. ROC curves illustrate that sensitivity and specificity are desirable ends that can not both be satisfied perfectly. The closest one can come to satiety (perfect sensitivity and specificity) is limited fundamentally by the inherent variances in the probability distributions of the two classes.

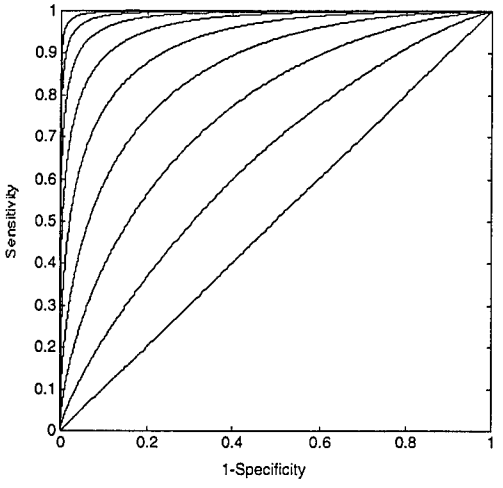


**Figure 11: Mathematical Family of ROC Curves**

To conform to the convention of drawing ROC curves that go from the lower left- to the upper right-hand corner, the horizontal axis features (1 − specificity) rather than specificity *per se*. The quantity plotted on the horizontal axis then shows the fraction of (actual) NEGATIVEs misidentified as POSITIVEs.

## D.3 Threshold Placement

Designing decisional algorithms is not as simple as merely selecting the class with the greatest membership probability. As an example, let us return to the two-class univariate problem, but such that the NEGATIVE and POSITIVE populations are highly asymmetric, with $\alpha_n = 97\%$ and $\alpha_p = 3\%$. These distributions are typical of many trauma registries, including the two that we analyzed in the present report.

Assuming, hypothetically, that maintaining equal specificity and sensitivity is a desired trauma management objective, the threshold, $\xi$, should always be set equal to $\frac{1}{2}(\mu_p - \mu_n)$. The specificity and sensitivity values in Eq. 4 are independent of $\alpha_p$ and $\alpha_n$. From Eq. 9, it follows that the required probability threshold is equal to

$$P_\xi = \frac{1}{1 + e^{-L_\xi}} \tag{5a}$$

in which

$$L_\xi = \theta_1 + \theta_x \xi = \ln(\alpha_n/\alpha_p) \tag{5b}$$

is logit value at the threshold point. It follows readily that $P_\xi = 0.03$. In other words, any patient with a nonsurvival probability above 3% should be declared a major trauma case and treated as such. The threshold placement closely matches the prevalence of major trauma among the general population that was used for fitting the underlying classification model. For this reason, special care must be taken to ensure that the trauma populations on which algorithms are trained closely resemble the populations to which they will later be applied.

## D.4 Reliability

A much more serious difficulty of working with asymmetric populations for training and testing algorithms is the poor *reliability* that results. Reliability indicators are a second set of ratios that follow from the confusion matrix, $\underline{\kappa}$, in which

$$R_{c,c'} \equiv \frac{\kappa_{c,c'}}{\sum_{c''} \kappa_{c,c''}} \tag{6}$$

is the percentage of exemplars assigned to class $c$ that actually belong to class $c'$. The denominator is computed by summing over rows, rather than columns, of $\underline{\kappa}$. The elements of the reliability matrix are related to those of the classification performance matrix through Bayes' formula, *viz.*

$$R_{c,c'} = \frac{\alpha_{c'} \pi_{c,c'}}{\sum_{c''} \alpha_{c''} \pi_{c,c''}}. \tag{7}$$

The rows of $R$ sum to unity. Fig. 12 illustrates a family of nominal reliability curves for the same set of $\Delta\mu/\sigma$ values as in Fig. 11. The curves are for the case of $\alpha_p = 0.03$. The true NEGATIVE rate, $R_{n,n}$, is plotted on the horizontal axis, with the true POSITIVE rate, $R_{p,p}$, on the vertical axis. Evidently, $R_{n,n}$ and $R_{p,p}$ are always greater than 0.97 and 0.03, regardless of the threshold placement. The latter quantity, $R_{p,p}$, tends to be quite poor when the population is highly asymmetric. For example, if $\Pi_{n,n} = \Pi_{p,p} = 0.95$, the true NEGATIVE and POSITIVE rates are 99.8% and 63.0% respectively. For discrimination powers of $\Pi_{n,n} = \Pi_{p,p} = 0.90$ and $\Pi_{n,n} = \Pi_{p,p} = 0.85$, the true POSITIVE rate falls to 21.8% and 14.9%. This phenomenon is in agreement with what was observed in the UVA and NCTR trauma data.

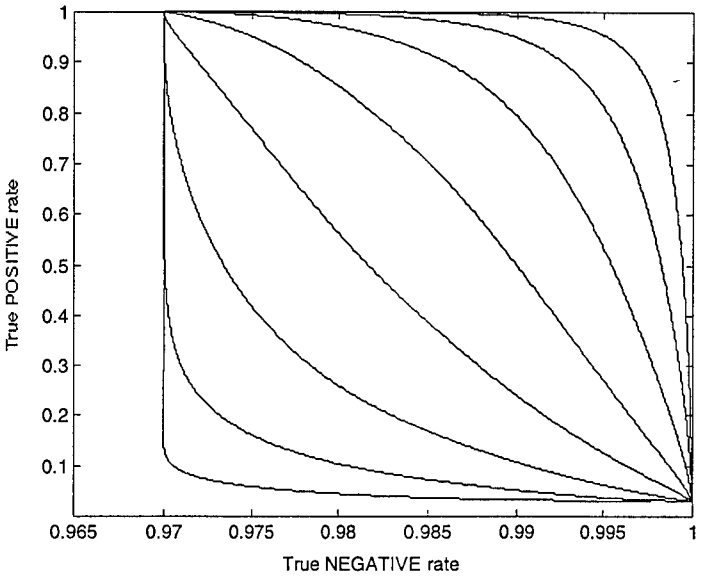


**Figure 12: Reliability Curves for $\alpha_p = 0.03$**

It is, of course, possible to improve the reliability statistics by changing the threshold placement, but this upsets the specificity and sensitivity characteristics. Moving the threshold in Fig. 10 to the right increases $R_{p,p}$ but reduces the sensitivity, resulting in undertriage. Where exactly the threshold should be properly placed, for a given level of discrimination capability between the classes, is one of the most challenging problems of trauma management. On the one hand, hospitals, insurance organizations, and public health officials, who are concerned primarily with *ex post* evaluations of critical care, are interested chiefly in sensitivity and specificity characteristics, i.e., $\Pi$. By contrast, EMTs in the field are concerned chiefly with whether pre-hospital decisions in individual cases are correct. They are interested mainly in the reliability statistics. If, for example, a hospital wishes to achieve performance criteria of 92% specificity and sensitivity, its EMTs would have to declare survival predictions that are actually erroneous three times out of four to avoid undertriage. Reconciling such conflicting objectives is incumbent on the medical community, but could possibly be resolved by analytic means through the systems-theoretic trauma management approach described in Appendix F.

# E Belief Networks

## E.1 Case Study

The most lucid way to illustrate belief network concepts is to work through a specific problem, such as the following case study introduced by Judea Pearl [63].

Consider a *state of affairs* defined by six *circumstances*:

$p_1$ = It is cloudy outdoors
$p_2$ = It is raining
$p_3$ = Your rain sensor tells you that it is raining outside
$p_4$ = Plans for a baseball game proceed
$p_5$ = Your son gets sunburned
$p_6$ = Your son goes off to visit his aunt.

Each circumstance is a proposition that is either true or false in the classical logic sense. However, you, the observer, cloistered inside your windowless office on a Saturday afternoon, lack complete factual information about the full state of affairs. By appealing to belief network formalism, however, you can obtain *circumstantial evidence* about the probability of your being able to attend the baseball game, even though you have no hard factual evidence about whether it will be raining outside.

The basic strategy is to construct a lookup table with one row for each possible state of affairs; in this case there would be $2^6 = 64$ rows. The first major step is to determine *a priori* probabilities for each possible state of affairs. This requires identifying cause-and-effect relationships among the various circumstances; in this case, everything depends, directly or indirectly, on weather it is cloudy outdoors. In the absence of any factual or circumstantial evidence about the whether, let us suppose that you can only conclude, from historical experience, that the probability that it is raining outside is 10%. From this, one would obtain the first of several *filter factors*, the product of which will give the *a priori* probabilities for the complete lookup table, *viz.*,

$$f_1 = 0.9 \cdot (\overline{p_1}) + 0.1 \cdot (p_1) \tag{1}$$

in which $\overline{p_1}$ denotes the statement that $p_1$ is false. Eq. 1 means that not knowing anything about how the presence of clouds affects any of the other circumstances, a state of affairs without clouds is nine times more likely than one with clouds. However, clouds determine whether it might be raining: if it is cloudy, there is, other things equal, a 60% chance that it will be raining outside, whereas it cannot possibly be raining on a cloudless day. This piece of general cause-and-effect knowledge generates a second filter factor, *viz.*,

$$f_2 = (\overline{p_1}) \cdot [1 \cdot (\overline{p_2}) + 0 \cdot (p_2)] + (p_1) \cdot [0.4 \cdot (\overline{p_2}) + 0.6 \cdot (p_2)] \tag{2}$$

Eq. 2 rules out entirely the state of affairs in which it is raining $(p_2)$ and cloudless $(\overline{p_1})$, since such is at variance with the certainty relationship that no clouds implies no rain.

Suppose that inside your windowless office, you have an unreliable rain alarm that has a 80% chance of sounding on a rainy day and a 4% chance of sounding on a fair day. This prompts a third filter factor:

$$f_3 = (\overline{p_2}) \cdot [0.96 \cdot (\overline{p_3}) + 0.04 \cdot (p_3)] + (p_2) \cdot [0.20 \cdot (\overline{p_3}) + 0.80 \cdot (p_3)] \tag{3}$$

If it is raining, there is a 95% chance that the game will be canceled, but if it is not raining, it is a certainty that the game will be held:

$$f_4 = (\overline{p_2}) \cdot [0 \cdot (\overline{p_4}) + 1 \cdot (p_4)] + (p_2) \cdot [0.95 \cdot (\overline{p_4}) + 0.05 \cdot (p_4)] \tag{4}$$

If it is cloudless, there is a 70% that your son will get sunburned, and a 10% that he will get sunburned on a cloudy day:

$$f_5 = (\overline{p_1}) \cdot [0.30 \cdot (\overline{p_5}) + 0.70 \cdot (p_5)] + (p_1) \cdot [0.90 \cdot (\overline{p_5}) + 0.10 \cdot (p_5)] \tag{5}$$

Finally, if your son is burned, he always runs off to his aunt for her to tend to him, but otherwise, there is only a 2% chance that he would have reason to see her:

$$f_6 = (\overline{p_5}) \cdot [0.98 \cdot (\overline{p_6}) + 0.02 \cdot (p_6)] + (p_5) \cdot [0 \cdot (\overline{p_6}) + 1 \cdot (p_6)] \tag{6}$$

Having covered all of the causal relationships among the six circumstances, the *a priori* probabilities may be computed directly, *viz.*,

$$P(p_1, p_2, p_3, p_4, p_5, p_6) = f_1 f_2 f_3 f_4 f_5 f_6 \tag{7}$$

This is a column of 64 numbers that sum to unity; once all of the causal relationships and filter factors have been accounted for, the resulting *a priori* probabilities are automatically normalized. The numbers in this example may be computed easily using almost any commercial spreadsheet package by following the steps described above. Of the 64, only 24 state-of-affairs scenarios have nonzero probabilities; the rest contradict certainty relationships. The scenario with the highest probability is that it is a cloudless, rainless day, your alarm does not sound, the game will proceed, your son gets burned, and he goes to his aunt. The *a priori* for this particular state of affairs is computed as:

$$P = (f_1 = 0.9) \cdot (f_2 = 1) \cdot (f_3 = 0.96) \cdot (f_4 = 1) \cdot (f_5 = 0.7) \cdot (f_6 = 1) = 0.6048 \tag{8}$$

In the absence of any partial evidence about the state of affairs on the particular day in question, this would be the most credible story. The *a priori* probabilities furnish classical conditional probabilities matching those probabilities that were invoked in the causal relationships. For example, the probability that it is raining given that it is cloudy is defined in the strict classical sense via combinatorics, *viz.*,

$$P(p_2|p_1) = \frac{P(p_2)}{P(p_1 \text{ and } p_2)} \tag{9}$$

The numerator is the sum of all *a priori* probabilities for which $p_2$ is true; the denominator sums over only those states for which $p_1$ and $p_2$ are both true. The result is 60%, which coincides with the probability used in the second filter factor. It is important to realize that only the *a priori* probabilities can be used in Eq. 9. Once you hear the alarm, it is *not* true that the credibility of rain-and-clouds divided by the credibility of rain is 60%.

The lookup table of *a priori* probabilities, once generated, furnishes a straightforward means of modifying the probabilities, or *belief weights*, of various combinations of circumstances as *partial* evidence is acquired. For example, suppose that your alarm does go off, and you wonder what this portends about the chances of your being able to attend the baseball game. In the absence

of any knowledge of the circumstances on the particular day, the lookup table would indicate that the game has a 94.3% chance of being held. Once you hear the alarm ring, however, you learn something peculiar to the state of affairs on the present day. Only those states for which $p_3$ is true can survive vis-à-vis the new evidence; all others have to be screened out. Of the 24 states with nonzero *a priori* probabilities, 12 are automatically eliminated in light of the new evidence. The *a priori* probabilities of the surviving 12 sum to 0.0856 . The *credibility* now ascribed to any one of those surviving scenarios is the *a priori* probability divided by 0.0856; this way, the credibilities sum to unity. It is by virtue of this renormalization process that circumstantial evidence can make some scenarios more credible than they previously were. For example, the scenario of clouds, rain, alarm, no game, no burn, no visit, which had an *a priori* credibility of 4.02%, now receives a credibility of 47.0%. It appears far more likely (53.3%) that the game will be canceled.

The screening and renormalization process would continue in this fashion as additional evidence is acquired. If, for example, your son telephones you to inform you of his intention to visit his aunt, without disclosing anything about the weather or his skin condition, the credibility of your being able to attend the game rebounds to 82.8%, since this would provide strong circumstantial evidence that your alarm had sounded falsely.

Belief networks can be applied to both classification, as in the case study just presented, and estimation. The only difference is that estimators perform reaveraging instead of renormalization. For example, suppose that the medical state of affairs for a trauma patient has been whittled down to one of four possible scenarios, the ISS scores corresponding to which are $10, 12, 14$, and 16. The estimated ISS score would therefore be $(10 + 12 + 14 + 16)/4 = 13$. If new evidence is acquired that contradicts the first two scenarios, the ISS belief would be revised to $(14 + 16)/2 = 15$.

## E.2 Belief Networks vs. Regression/PNN Methods

Regression and PNN methods for estimation and classification relate inputs and outputs by way of explicit *functions*, i.e., $y = f(\underline{X})$. Belief networks, by contrast, take the radically different approach of obtaining output estimates and probabilities from lookup tables. To appreciate the basic difference in the regression/PNN and belief network approaches, it is perhaps best to think of the two conceptually in the following manner. Regression and PNN methods attempt to fit a single function across the entirety of the input variable space, i.e., a continuous, real-valued function in the case of estimators and such functions as logits in the case of classifiers. The validation process, governed by PSE or cross-validation, generally forces these functional forms to be parsimonious to avoid overfit. Prevention of overfit mandates that the fitted functions, even high-degree Ivakhnenko polynomials, not have excessive curvature. As a result, fitted models tend to be "stiff," a linear regression, in the most rudimentary case, forcing a perfectly straight line through a set of data points.

A major disadvantage of regression/PNN is that these methods attempt to *summarize* the entirety of training data presented to them. Their one-function-fits-all paradigm tends to make it difficult for them to adapt to local peculiarities in certain regions of input space. Alternatively interpreted, the fitted functions cannot have locally rough or sharp features. However, whereas regression models are irrevocably handicap with respect to such objections, PNNs do have some flexibility to accommodate local adaptation: (1) they admit more general (and therefore higher curvature) functions than regression models; (2) locally trained PNN models can be "spliced" together in a mathematically clean fashion; and (3) the PNN method can readily use nonpolynomial

basis functions instead. One particular genre of admissible basis functions has been formulated chiefly to address this type of dilemma. In signal processing applications, wavelets have been introduced as special basis functions that have translational and dilatational degrees of freedom. This enables them to zoom in on localized "blips," e.g., seismographic murmurs, musical notes, speech sounds, optical images, electrocardiogram anomalies. Because of this flexibility, wavelets are able to overcome the well-known limitations of Fourier transform methods arising from the uncertainty principle. There is, in a vague sense, a "stiffness" in the Fourier basis functions analogous to that of the polynomial functions used in regression.

An alternative, somewhat less elegant, method for addressing the local adaptation problem is simply to chop the input space up into a collection of boxes or cells. A crude, but clever, example would be to treat all trauma patients with GCS = 15 and normal RR as one cell and others with GCS = 14, normal RR as another, completely independent cell. The estimated ISS for any patient in a given cell would be the mean ISS historically observed for all past patients having belonged to that cell. This is exactly the belief network approach, in which each "cell" is a row of the lookup table. In this way, belief networks, which treat each scenario cell individually and independent of all others, decouple the various regions of input space and overcome the one-function-fits-all drawbacks of regression models. Although causal considerations may have been used to compute *a priori* probabilities, the point is that such steps are not necessary to compute them; the *a priori* probabilities in the Appendix E case study could just as well have been chosen randomly, filled in by hand, and normalized. Conversely, if one obtained a complete table of *a priori* probabilities empirically, it would be possible, through inspection of the numbers, to deduce the nature of the causal relationships among the circumstances. This, however, is a complicated algorithmic process outside the present scope.

Belief networks also provide a natural framework for *graceful degradation* of models, wherein certain input data fields are omitted. In, for example, a belief network to estimate ISS from GCS and RR, one would construct a lookup table of length equal to the number of GCS bins times the number of RR bins. If GCS and RR are both available, the ISS would be looked up directly. But if, say, RR were not available, the best one could do to infer ISS would be to average ISS over all rows whose GCS matches the known value. In a regression strategy, by contrast, one would have to do something cruder, such as develop a separate GCS-only backup model from scratch or assume an average value for GCS obtained from the training database.

The advantages of regression methods over belief networks is primarily one of synthesis speed and ease of modeling. Furthermore, unlike belief networks, regression models correctly recognize that there generally should be some smoothness and continuity in the outputs generated by neighboring cells. Whereas the screening and renormalization/reaveraging processes for belief networks tend to be arduous and cumbersome, computing a polynomial function is a snap. On the other hand, the difficult and lengthy validation process for regression models is avoided entirely in belief networks by virtue of the single-cell paradigm.

# F Systems-Theoretic Approach to Trauma Management

Having described in Section 2 the environments in which trauma care takes place, we now elaborate further the conceptual scope of trauma management. To help make a case for how triage algorithms will fit into the big picture, we present a general systems-theoretic framework for interpreting trauma management in its entirety as an integrated process in which algorithmic software-driven tools would play a powerful and decisive role. Consideration of the intricate and profound aspects of the larger problem, we believe, is ultimately necessary to overcome the most difficult impasses in developing effective and accurate algorithms for injury severity and outcome prediction.

## F.1 Multidisciplinary Nature of Trauma Management

The problem of trauma management, at the highest level, is multidisciplinary. On the one hand, its focus is the human body. Every trauma care decision, both in the field and in the hospital, depends on how the patient's bodily condition is expected to evolve in the near future, assuming the application of certain treatment. Prediction of physiological outcomes, chiefly life and death, is the most fundamental challenge at the heart of trauma management. It is also by far the most difficult and involved part of the puzzle. The human body is a highly complex system of anatomical structures, cells, and biochemical processes all tightly interacting in a purposeful, coherent, and well-regulated fashion. Determining how it responds to a specific type of traumatic disturbance is clearly a problem of human physiology [35] and closely related disciplines in the biological sciences.

Another major component of trauma management, namely the design and development of biomedical instrumentation to acquire medical data from patients, also has roots in the biological sciences. Virtually every remaining aspect of the problem, on the other hand, draws on disciplines in the quantitative sciences, e.g., queueing theory, modeling, biostatistics, pattern classification, and decisional theory. These fields belong to the more general domains of applied mathematics, systems engineering, and management science. Collectively, all of these diverse tools must be harnessed in a concerted effort to make the most of what limited biological data on patients can feasibly be acquired.

The biological and mathematical aspects of the general trauma management problem can be decoupled in the following sense. Suppose that the physiology of the human body and trauma (Appendix B) were so completely understood that the fate of a patient having sustained a given precisely-defined type of injury could be ascertained deterministically. Imagine a computer simulation of the complete physiological condition of the patient as a dynamical trajectory through time. In the language of dynamical systems and control theory, the body, in the absence of medical treatment, would be regarded as an autonomous dynamical system, or plant, and medical intervention would be modeled as a set of exogenous control inputs, or forces, applied to the system. Observable outputs would correspond to various clinical indicators, such as blood pressures and respiratory status, that EMTs could readily obtain either by direct perception or with biomedical instruments. The entire biological component of the problem could then be captured fully by an integrated software suite that would: (1) generate a distribution of injury incidence patterns (i.e., initial conditions) appropriately characterizing a particular civilian or military environment; (2) simulate the evolution of the patient's medical condition through time in full detail; (3) continuously accept medical treatment input and respond thereto; and (4) continuously supply output

data. With such a hypothetical blackbox simulation tool, the quantitative science methodologies, as mentioned above, furnish the full arsenal of tools and scientific resources needed to *solve* the trauma management problem. In other words, they can be applied to produce a protocol of trauma care *policies* such that if EMTs undertake certain prescribed treatment actions in response to a stream of output data for a given patient, loss of life and limb will be minimized. Since scarcity of treatment resources, as well as purely medical criteria, must be considered, the overall measure of trauma management effectiveness would be minimizing aggregate losses.

Even with such a fantastic (at least by contemporary standards) blackbox simulator, the quantitative-science part of the trauma problem remains formidable and involved. Suppose, for example, that given any injury case, it were possible to determine immediately and with complete accuracy the type and intensity of emergency medical care required to save the patient. Given that trauma care resources are scarce and finite lead times (e.g., evacuation transit, cross-matching of blood types) antecede actual arrival of some life-saving services, determining which patients to attempt to save and where to send them is a queueing problem, which, in its own right, is far from trivial. Moreover, the medical conditions of patients deteriorate as they "wait" in the queue. Queueing simulations would be needed to optimize the triage and patient selection rules. To complicate matters, appropriate treatment and, thus, the necessary bundle of life-saving resources are not known *a priori*. The EMT and critical care hospital surgeon have wide choices of treatment options, some of which might prove counterproductive. The consequences of each alternative would have to be explored. Even if the EMT or physician had complete knowledge of the physiological processes inside a patient's body, it would be a doubly complicated problem to optimize the treatment and queueing rules simultaneously. To make matters worse yet, the totality of sensors that the EMT could expect to have, given current technology, could not come anywhere close to providing an exhaustive window of knowledge as to what is happening inside the patient's body. Very much to the contrary, short lists of salient features such as systolic blood pressure (SBP), Glasgow Coma Scale (GCS), and respiratory rate (RR) provide only a woefully sparse and often inconclusive glimpse into underlying physiological processes. Such data can only provide a portrait of the patient's condition so sketchy that potential outcomes must be treated probabilistically, with great overlap of the attributes of surviving and nonsurviving patients. The highly stochastic nature of the problem, due to the limitations on what aspects of the physiological processes can be *observed*, greatly complicates the formulation of treatment and queueing rules.

Nevertheless, the problem is still, in principle, amenable to solution. The decoupling approach is intriguing in that it promises a full-blown solution to the entire trauma management problem. Moreover, all of the biological elements could be contained in the blackbox simulator. The quantitative-science part of the problem, as just discussed, could be solved using resources no more than pencil-and-paper analysis, computing power (with extremely high speeds and storage), and knowledge of the trauma environment for purposes of building realistic parameters and assumptions into the simulations.

## F.2 Integration of Simulation Methods into Trauma Management

To devise useful quantitative tools to drive trauma treatment, it is clearly necessary to have some means of testing their performance. Simulation, in principle, is a powerful methodology well suited for applications such as this, in which it is desirable to trace the state trajectory of a complex dynamical system such as the human body. The blackbox simulator that we have envisioned, for

instance, would be able to account for all of the various physiological effects in response to any given injury. The ability to perform such simulations, once they reach a certain level of sophistication, will undoubtedly play an invaluable role in helping solve the entire trauma management problem. Most significantly, it would breach the most vexing impediment that all efforts to date have encountered, namely the dearth of original trauma data. Clearly, it is not possible to go out and obtain empirical data at will; exemplars can only be obtained with hindsight from actual trauma cases that were unfortunate enough to have occurred.

There is thus an extremely compelling demand for software methods to simulate trauma in virtual human subjects. Such an approach would have at least four remarkable advantages: (1) they could accurately account for most of the understood physiological aspects of response to trauma in humans; (2) simulation runs could be performed in arbitrarily copious numbers; (3) individual simulations could be rerun to explore the consequences of different medical intervention alternatives at various times; and (4) they are nondestructive and do not require experimental tests. Moreover, they completely isolate the biological aspects of the trauma management problem. For these reasons, this type of simulation capability, once the state of knowledge and software technology to realize it becomes available, will make the trauma management a far more tractable problem than it is today.

Trauma simulation could be utilized in developing models and protocols for trauma management. We now clarify the meaning of *trauma management* and highlight the types of decisions that need to be made during the process. In doing so, we focus on the pre-hospital triage elements of trauma management, i.e., those decisions about the qualitative degree of care required by a patient, based on: (1) the limited biological information about his/her condition revealed through clinical indicators; and (2) consideration of treatment resource scarcity. As our own contribution of expertise, we have shown how to develop, interpret, and validate *algorithms* to drive triage, i.e., mathematically clear-cut procedures to dictate decisions, based on all of the various biological output indicators (which would be supplied as *inputs* to the algorithms).

From a systems-theoretic perspective, trauma management may be viewed as a set of sequential processes, from the traumatic event itself to discharge from a *trauma center*, meaning any institution dedicated to providing emergency care to critical patients. Trauma centers may include highly specialized critical care units in hospitals, general-purpose emergency rooms, or makeshift treatment centers (during war or in impoverished societies). In any such institutional setting, an incoming patient stands to receive at least a semi-professional level of medical care that could make a difference between life and death. Abstractly, the trauma center may be viewed as a set of parallel *servers* catering to a queue of incoming patients. A server, in this view, is a bundle of reusable resource fixtures (e.g., beds, teams of medical personnel) that a patient may require during his stay in the trauma center. To simplify the modeling effort, various assumptions may be made (e.g., that all nonreusable resources, such as blood for transfusions, are infinitely abundant, that a patient is under the undivided medical attention of a single server until discharged from the center, and that all servers are equivalent and capable of performing any technologically possible medical procedure that may be appropriate). As soon as a server becomes vacant, it immediately accepts a patient (provided that at least one exists) from the queue. In receiving a new patient, the server would consult a *selection algorithm* to identify that patient in the queue in most critical need of immediate attention, based on pre-hospital information about the patients' conditions. Based on such information, the blackbox simulator could reveal what would happen to each patient in the queue assuming either immediate attention or further wait. It could also indicate under what

circumstances it becomes safe and appropriate to discharge a patient.

Trauma centers, as integral units in the larger trauma management system, could thus optimize their own operating policies by utilizing selection algorithms in receiving patients for treatment. The detailed modeling of the trauma center, of course, could be modified to reflect more realistically how such institutions actually operate. For example, the servers may not all be equivalent in terms of the quality of care they can provide, and different institutions (e.g., Level I and Level II hospitals) may generally be expected to have differing grades of servers. Based solely on the demand patterns (i.e., mean arrival rates and distributions of injury types encountered), any given trauma center could perform simulations to determine optimal selection rules and treatment procedures for its own internal use. Such queueing concepts, it should be pointed out, are of great practical interest today insofar as medical information on pre-hospital patient status, forwarded by EMTs en route to hospital, enables hospital emergency departments to prepare for such imminent arrivals and to begin appropriate treatment immediately. Pertinent physiological data, such as those mentioned in the solicitation and in Appendix B, enable hospital physicians to construct *risk stratification* profiles of incoming patients and manage institutional resources more effectively in response to demand patterns.

The trauma center end of the problem having been thus solved, pre-hospital EMTs could determine the probability of a given patient under their care being treated successfully by any one of several alternative trauma center destinations. Given knowledge of the patient's medical condition, the pre-hospital EMT has a number of *evacuation modality* and destination options. For a given evacuation modality (e.g., land ambulance, aeromedical transport), there will be a certain transit time that can usually be predicted quite reliably. Upon arrival of the patient at a given trauma center, the EMT may anticipate a certain probability distribution describing the waiting period that patient would face, based on the actual demand or the demand pattern that the hospital usually experiences, and its selection rule policies. The triage decision (choice of evacuation modality and destination) reflects both the severity of the patient's medical condition and consideration of delay times. Triage algorithms, as part of a fully integrated solution to the trauma management problem, would have to be able to assume such decisional burdens to aid pre-hospital EMTs. Separate *treatment directive* algorithms would determine the most appropriate short-term treatment procedure, based solely on medical information about the patient's condition.

## F.3 Future Role of Trauma Simulation

The discussion and exposition in the preceding sections have sought to portray the true complexity of the scientific challenges at hand and a vision of the larger trauma management problem into which triage algorithms, we believe, properly fit. We have presented the dimensions of the larger problem and the intricate nature of decisions that need to be made both in the field and in the hospital.

The biological part of the trauma management problem is by far the most difficult. Simulating the detailed physiological response of the body to a certain type of traumatic injury is presently not practical. In the future, it is certainly conceivable that such capability will be realized at the level of detail, simulation speed, and sophistication that would be needed for the purposes we have described. We believe that it will be the ultimate key to a truly momentous breakthrough in trauma management, and that once realized, will enable the entire problem to be tackled in principle. A vast amount of research in the area of physiological modeling has already been done

(cf., the ARPA biomedical project called "MediSim: Simulated Medical Corpsmen and Casualties for Medical Forces Planning and Training" being performed in conjunction with the Medical College of Pennsylvania, Sandia National Laboratory, and the Naval Postgraduate School), and much could be accomplished by using neural networks to provide blackbox models where analytic subsystem models are not presently available. However, we do not discuss physiological modeling further in either the present Phase I Final Technical Report or in our Phase II proposal.

The present state of technology forces reliance on historical data from actual trauma incidents. Such databases, however, have inherent drawbacks that must be acknowledged upfront. They seldom contain patient records in the large numbers needed to perform truly conclusive statistical analyses. The difficulty in acquiring access to civilian trauma registries for studies such as this reflects, in part, concerns by the owners of such databases that inter-hospital comparisons may be "unfair." Opportunities to collect military trauma data, in particular, are extremely rare; the only noteworthy example of such a body of data of which we are aware is the Wound Data and Munitions Effectiveness in Vietnam (WDMEV) database [9]. Despite severe limitations such as these, much useful analysis can be done on trauma registry data, primarily because algorithms can, and must, play a key role in both the very nonideal world of today and the much-closer-to-ideal world (in which simulation capability is readily available) of tomorrow. For this reason, demonstrating how to harness algorithms, even on contemporary data, is by no means a moot exercise. Understanding the limitations of conventional trauma registry data, however, is essential for sound statistical analysis and knowing where problems and weaknesses lie.

# G Conventional Nonlinear Regression Approach

## G.1 Modeling Using Nonlinear Stepwise Regression

In conventional *nonlinear regression* approaches to estimation problems, $f$ is usually taken to have an algebraic polynomial functional form, *viz.*,

$$y \approx \hat{y} \equiv \underline{\theta}^T \cdot \underline{x} \tag{10}$$

in which $\underline{x}$ is a *synthetic* column vector containing monomial product combinations of the raw inputs, $\underline{X}$, and $\underline{\theta}$ is the corresponding set of multiplicative coefficients (model parameters). For example,

$$\underline{x} = [1 \quad \text{GCS}^2 \quad \text{GCS}\cdot\text{RR}]^T$$

is a set of synthetic inputs constructed from the raw input set $\underline{X} = [\text{GCS} \quad \text{RR}]$. Note that the inner product expression $\underline{\theta}^T \cdot \underline{x}$ in this way represents a general polynomial function. The analogous procedure for logistic regression is to generalize the logit polynomials (Appendix C.3) in the same way via synthetic inputs.

In this appendix, we illustrate the steps typically required for structure learning without use of an ontogenic neural network synthesis tool such as *GNOSIS*, for purposes of training and validating regression models; the trauma data from the University of Virginia are used in this example. The purpose here is merely to illustrate the key steps of conventional nonlinear stepwise regression and to contrast the approach with neural network synthesis algorithms, which automatically achieve these basic objectives while yielding superior models. We herein illustrate, by example, all of the key steps involved in constructing and validating regression models to provide values of the output (ISS) as an explicit function of the inputs.

For a given polynomial model structure, the coefficient values are computed readily via the least-squares algorithm. Choice of model structure, however, is an open question left entirely to the discretion of the analyst. We start with only a vague notion that ISS may somehow be related functionally to the five key input variables provided in the database (AGE, B/P, GCS, RR, SBP). A systematic search of candidate structures is needed. One basic stepwise regression strategy is to start with a large quadratic polynomial and remove, or carve away, unnecessary terms one-by-one. Once an optimally lean quadratic model is found, the process is repeated starting with cubic and higher degree polynomials.

For illustration purposes, we start with a model containing a constant term, B/P as a linear term, plus a complete quadratic polynomial in GCS, RR, and SBP. Let us denote this structure as 1-$b$-$g$-$r$-$s$-$g^2$-$gr$-$gs$-$r^2$-$rs$-$s^2$, in which the lowercase letters are abbreviated mnemonics for the input variable and '1' denotes the constant term. This is a very liberal, unparsimonious structure which probably overfits the data. To test it, we use not the entire set of exemplars for training but only part of it (e.g., 70%). The remainder of the dataset (e.g., 30%) is reserved for evaluation of estimation errors. The purpose of partitioning the database this way is to account for the fact that no training database, no matter how extensive, can include unforeseen cases that have not yet been encountered. A real pre-hospital triage algorithm, for instance, would contend in the field with individual cases that obviously were not included in the original training database *per se*. This hardship can be addressed by training on a truncated database and evaluating performance with exemplars on which the model was not trained. When the entire database is sparce small, the partition cut must be very shallow to avoid making the training databases too small. The

procedure commonly followed in such cases, known as *jackknifing*, withholds just one exemplar for evaluation and trains on all others. The procedure is repeated for all exemplars in the database. For a given model structure, a distribution of values for each coefficient slot is obtained.

The procedure for large databases is essentially the same, except that significantly deeper cuts can generally be made. The process is repeated with different random cuts (all of the same depth); 100 such repeats is often reasonable. In each cut, the resulting model coefficient values are computed; these vary depending on which exemplars are randomly assigned to training database. The estimation errors are then calculated for each exemplar in the complementary evaluation database. Over the evaluation database as a whole, this furnishes a distribution of estimation errors having a mean and standard deviation. Whereas the mean of the estimation-error distribution is typically close to zero, the standard deviation is often appreciable. This standard deviation statistic, averaged over the 100 random cuts, is generally a stable quantity that serves conveniently as a benchmark index for the *performance* of a proposed model structure. For $1\text{-}b\text{-}g\text{-}r\text{-}s\text{-}g^2\text{-}gr\text{-}gs\text{-}r^2\text{-}rs\text{-}s^2$, this figure of merit was approximately 6.58. This means that with this model structure, one can expect to be in error by roughly this amount in field estimates of ISS. There may be three ways to obtain better results:

- Focus on *exemplar quality*. The field measurements themselves provided in the database may be uncertain or inaccurate.
- Focus on *database comprehensiveness*. The database may be too small or sparse to capture representative manifestations of trauma. Alternatively interpreted, the graph of a function cannot be resolved or recognized with too few plotted points.
- Develop new biomedical instrumentation to obtain additional data fields that may improve medical assessment of the patient.
- Try different candidate model structures.

either add or drop terms; both are double-edged swords. Dropping terms may throw away valuable information that the existing model has captured already. Adding terms runs the risk of overfit. Modeling data using a high-order polynomial may work nicely over a limited region, for instance, but would result in poor performance if applied to unseen data far removed from that region, since the high-order terms diverge rapidly.

Because of the danger of overfit and the extra computational burden thereby introduced, it is generally preferable to drop rather than add them; in other words, to work down toward a more parsimonious structure. This requires a method of identifying those terms that contribute least to the existing model and can therefore be omitted most prudently. One way to do this is to examine the distribution of coefficient values for each term over repeated cuts. The means and standard deviations, for a particular 100-cut trial, are tabulated in Table 36. The relevance of each variable may be assessed summarily by computing its *coefficient of variation*, or the ratio of the standard deviation to the mean. The constant term, for example, has a COV of 0.09, which indicates that the value of this coefficient reliably lies between 21 and 26 most of the time. The $rs$ term, by contrast, has a COV of 3.9, which indicates that the value of this coefficient is highly erratic and unpredictable, assuming wildly varying positive and negative values in different cuts. The significance of the $rs$ term is highly ambiguous and unclear; it is therefore reasonable to drop it. The overall *leanness* of the model may be judged by the COVs of its coefficients, and a leanness figure of merit may be formally defined as the maximum of the set of COVs. A lower leanness score indicates a leaner model than a high leanness score.

**Table 36: Model Coefficients for Unparsimonious Model**

| | |
|---|---|
| 1 | $23.7 \pm 2.2$ |
| $b$ | $-1.41 \pm 0.24$ |
| $g$ | $1.63 \pm 0.44$ |
| $r$ | $0.306 \pm 0.109$ |
| $s$ | $-0.0767 \pm 0.0208$ |
| $g^2$ | $-0.196 \pm 0.024$ |
| $gr$ | $0.0078 \pm 0.0075$ |
| $gs$ | $0.0015 \pm 0.0011$ |
| $r^2$ | $-0.0028 \pm 0.0014$ |
| $rs$ | $0.0002 \pm 0.0006$ |
| $s^2$ | $0.0002 \pm 0.0001$ |

Structure learning proceeds by discarding the term with the highest COV until a model structure representing the best compromise between performance and leanness is found. Results are tabulated in Table 37. The performance index, evidently, is extremely difficult to drive down; the table shows

**Table 37: Performance and Leanness Indices for Alternative Model Structures**

| Structure | Performance | Leanness |
|---|---|---|
| 1-$b$-$g$-$r$-$s$-$g^2$-$gr$-$gs$-$r^2$-$rs$-$s^2$ | 6.5787 | 3.91 |
| 1-$b$-$g$-$r$-$s$-$g^2$-$gr$-$gs$-$r^2$-$s^2$ | 6.5926 | 1.01 |
| 1-$b$-$g$-$r$-$s$-$g^2$-$gs$-$r^2$-$s^2$ | 6.5904 | 1.13 |
| 1-$b$-$g$-$r$-$s$-$g^2$-$gs$-$s^2$ | 6.5421 | 0.81 |
| 1-$b$-$g$-$r$-$s$-$g^2$-$s^2$ | 6.5081 | 0.23 |
| 1-$b$-$r$-$s$-$g^2$-$s^2$ | 6.5596 | 0.33 |
| 1-$b$-$r$-$s$-$g^2$ | 6.5888 | 0.50 |

only a 1% gain. The leanness score, on the other hand, is reduced substantially: down by a factor of 17 from 3.91, for the most liberal structure, to 0.23 for 1-$b$-$g$-$r$-$s$-$g^2$-$s^2$. This model also has the best performance index and would therefore be the most appropriate structure to exploit. The coefficients for this structure are tabulated in Table 38. The means and standard deviations of the coefficients are similar to those in the more liberal structure first surmised. The unnecessary terms, however, have been carved away.

## G.2 Advantages of *GNOSIS* over Regression

We next discuss the key steps involved in applying *GNOSIS* to dramatize the tremendous labor-saving advantages over least-squares and logistic regression that it offers to the analyst.

In synthesizing estimation models from a training database, *GNOSIS* uses default settings of three inputs per node, nodal outputs that are cubic polynomial functions of the nodal inputs, and a maximum of four layers. Layers are synthesized sequentially. The least valuable nodes

**Table 38: Model Coefficients for Reduced Model**

| | |
|---|---|
| 1 | $22.3 \pm 1.7$ |
| $b$ | $-1.43 \pm 0.24$ |
| $g$ | $1.91 \pm 0.42$ |
| $r$ | $0.249 \pm 0.024$ |
| $s$ | $-0.0695 \pm 0.0166$ |
| $g^2$ | $-0.193 \pm 0.022$ |
| $s^2$ | $10^{-4} \times (2.30 \pm 0.58)$ |

are automatically carved away. Once the nodes in a given layer have been synthesized, *GNOSIS* can further refine the layer by creating additional nodes whose inputs are not only outputs from the previous layer but also outputs from the just-generated nodes within the current layer. This technique, known as *projection pursuit*, substantially enhances the performance of the resulting PNN model.

**Table 39: *GNOSIS* Performance with and without Projection Pursuit**

| Layer number | RMS Estimation Error without projection pursuit | RMS Estimation Error with projection pursuit |
|---|---|---|
| 1 | 6.401 | 6.322 |
| 2 | 6.325 | 6.246 |
| 3 | 6.285 | 6.200 |
| 4 | 6.265 | 6.190 |

Table 39 displays the root-mean-square (RMS) estimation error from *GNOSIS*-synthesized models (with default settings) obtained after successive layers have been completed. The numbers show that each additional layer furnishes a more accurate model than the previous layer. For example, a four-layer model without projection pursuit provides a RMS estimation error of 6.265, which is appreciably better than 6.401 for a single-layer model (which is just a cubic regression). The gains in estimation error reduction, however, diminish as the number of layers increases. After a certain number of layers, no further modeling improvement is realized.

Improvement over the results in Table 39 can be achieved by using non-default settings. For example, the number of nodal inputs could be increased from the default setting of three to four, in which case the projection-pursuit RMS estimation errors in the first and second layers are reduced to 6.224 and 6.103 respectively. This represents a significant gain in accuracy *without overfit.* However, the number of input permutations is so much larger than in the three-input default that the synthesis process takes considerably longer. A much faster and even more effective way to reduce error is to admit fourth- or fifth-degree nodal polynomials. The resulting performance gains are displayed in Table 40.

**Table 40: *GNOSIS* Performance with High-Order Nodal Polynomials**

| Layer number | RMS Estimation Error 3rd-degree nodes | RMS Estimation Error 4th-degree nodes | RMS Estimation Error 5th-degree nodes |
|---|---|---|---|
| 1 | 6.322 | 6.227 | 6.162 |
| 2 | 6.246 | 6.042 | 6.969 |
| 3 | 6.200 | 5.984 | 5.892 |
| 4 | 6.190 | 5.965 | 5.848 |

All of the models in Table 40 use projection pursuit. Clearly, performance greatly improves with higher-degree nodal polynomials. Utilization of such polynomials is not as costly in synthesis time as is allowing additional nodal inputs; it merely means that the least-squares fitting of coefficients at each node involves more degrees of freedom. However, the synthesis times for fifth-order polynomials are sufficiently inconvenient that we chose to rely on fourth-degree models as the basis for definitive results documented in this report.

The best result, an RMS estimation error of 5.848 for fifth-order polynomial nodes and four layers, represents a 12% reduction in the RMS estimation errors for least-squares regression models. A lengthy and convoluted search and structure comparison process using a conventional regression approach was unable to bring the RMS estimation error below 6.47. All of the effort to cascade down to the "best" structure thus led to a low-order polynomial that, in comparison to the PNN results, is very poor indeed. Moreover, the much better scores accruing to the PNN models were obtained with significantly less computational effort, in terms of both machine computations and burdens imposed on the analyst. All that the *GNOSIS* user need do, essentially, is stipulate the degree of the nodal polynomials and watch the error statistic diminish as successive layers are generated. In regression modeling, by contrast, the analyst literally needs to catalogue many conceivable model structure and test them (with cross-validation) one-by-one. The process is prohibitively arduous, whether done via heuristic inspection (as demonstrated in the preceding discussion) or via automated stepwise regression algorithms. Owing to AIC and PSE, however, the analyst does not even have to perform time-consuming cross-validations and, in principle, is free to work with the entire training database for each test with a given choice of settings.

## Personnel Receiving Pay from This Effort

George Lindbeck, M.D.
Edward C. Larson
B. Eugene Parker, Ph.D.
Susan L. Woodson
Lori P. Seal
Roger L. Barron
David P. Wipf
R. Dale Salmons

**DEPARTMENT OF THE ARMY**
US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND
504 SCOTT STREET
FORT DETRICK, MARYLAND 21702-5012

REPLY TO
ATTENTION OF:

MCMR-RMI-S (70-1y) 4 Dec 02

MEMORANDUM FOR Administrator, Defense Technical Information Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir, VA 22060-6218

SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to technical reports written for this Command. Request the limited distribution statement for the enclosed accession numbers be changed to "Approved for public release; distribution unlimited." These reports should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Kristin Morrow at DSN 343-7327 or by e-mail at Kristin.Morrow@det.amedd.army.mil.

FOR THE COMMANDER:

Encl

PHYLIS M. RINEHART
Deputy Chief of Staff for
Information Management

ADB218773
ADB223531
ADB230017
ADB223528
ADB231930
ADB226038
ADB224296
ADB228898
ADB216077
ADB218568
ADB216713
ADB216627
ADB215717
ADB218709
ADB216942
ADB216071
ADB215736
ADB216715
ADB215485
ADB215487
ADB220304
ADB215719
ADB216072
ADB222892
ADB215914
ADB222994
ADB216066
ADB217309
ADB216726
ADB216947
ADB227451
ADB229334
ADB228982
ADB227216
ADB224877
ADB224876
ADB227768
ADB228161
ADB229442
ADB230946
ADB230047
ADB225895
ADB229467
ADB224342
ADB230950
ADB227185
ADB231856
ADB229914
ADB229497
ADB230947
ADB282209
ADB270846
ADB282266
ADB262442
ADB256670